

KINASES AND PHOSPHATASES

TECHNICAL FIELD

The invention relates to novel nucleic acids, kinases and phosphatases encoded by these
5 nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and kinases and phosphatases.

10

BACKGROUND OF THE INVENTION

Reversible protein phosphorylation is the ubiquitous strategy used to control many of the intracellular events in eukaryotic cells. It is estimated that more than ten percent of proteins active in a typical mammalian cell are phosphorylated. Kinases catalyze the transfer of high-energy phosphate
15 groups from adenosine triphosphate (ATP) to target proteins on the hydroxyamino acid residues serine, threonine, or tyrosine. Phosphatases, in contrast, remove these phosphate groups.

Extracellular signals including hormones, neurotransmitters, and growth and differentiation factors can activate kinases, which can occur as cell surface receptors or as the activator of the final effector protein, as well as other locations along the signal transduction pathway. Cascades of kinases occur,
20 as well as kinases sensitive to second messenger molecules. This system allows for the amplification of weak signals (low abundance growth factor molecules, for example), as well as the synthesis of many weak signals into an all-or-nothing response. Phosphatases, then, are essential in determining the extent of phosphorylation in the cell and, together with kinases, regulate key cellular processes such as metabolic enzyme activity, proliferation, cell growth and differentiation, cell adhesion, and
25 cell cycle progression.

KINASES

Kinases comprise the largest known enzyme superfamily and vary widely in their target molecules. Kinases catalyze the transfer of high energy phosphate groups from a phosphate donor to a phosphate acceptor. Nucleotides usually serve as the phosphate donor in these reactions, with most
30 kinases utilizing adenosine triphosphate (ATP). The phosphate acceptor can be any of a variety of molecules, including nucleosides, nucleotides, lipids, carbohydrates, and proteins. Proteins are phosphorylated on hydroxyamino acids. Addition of a phosphate group alters the local charge on the acceptor molecule, causing internal conformational changes and potentially influencing intermolecular contacts. Reversible protein phosphorylation is the primary method for regulating
35 protein activity in eukaryotic cells. In general, proteins are activated by phosphorylation in response

to extracellular signals such as hormones, neurotransmitters, and growth and differentiation factors. The activated proteins initiate the cell's intracellular response by way of intracellular signaling pathways and second messenger molecules such as cyclic nucleotides, calcium-calmodulin, inositol, and various mitogens, that regulate protein phosphorylation.

5 Kinases are involved in all aspects of a cell's function, from basic metabolic processes, such as glycolysis, to cell-cycle regulation, differentiation, and communication with the extracellular environment through signal transduction cascades. Inappropriate phosphorylation of proteins in cells has been linked to changes in cell cycle progression and cell differentiation. Changes in the cell cycle have been linked to induction of apoptosis or cancer. Changes in cell differentiation have been linked
10 to diseases and disorders of the reproductive system, immune system, and skeletal muscle.

There are two classes of protein kinases. One class, protein tyrosine kinases (PTKs), phosphorylates tyrosine residues, and the other class, protein serine/threonine kinases (STKs), phosphorylates serine and threonine residues. Some PTKs and STKs possess structural characteristics of both families and have dual specificity for both tyrosine and serine/threonine
15 residues. Almost all kinases contain a conserved 250-300 amino acid catalytic domain containing specific residues and sequence motifs characteristic of the kinase family. The protein kinase catalytic domain can be further divided into 11 subdomains. N-terminal subdomains I-IV fold into a two-lobed structure which binds and orients the ATP donor molecule, and subdomain V spans the two lobes. C-terminal subdomains VI-XI bind the protein substrate and transfer the gamma phosphate from ATP to
20 the hydroxyl group of a tyrosine, serine, or threonine residue. Each of the 11 subdomains contains specific catalytic residues or amino acid motifs characteristic of that subdomain. For example, subdomain I contains an 8-amino acid glycine-rich ATP binding consensus motif, subdomain II contains a critical lysine residue required for maximal catalytic activity, and subdomains VI through IX comprise the highly conserved catalytic core. PTKs and STKs also contain distinct sequence
25 motifs in subdomains VI and VIII which may confer hydroxyamino acid specificity.

In addition, kinases may also be classified by additional amino acid sequences, generally between 5 and 100 residues, which either flank or occur within the kinase domain. These additional amino acid sequences regulate kinase activity and determine substrate specificity. (Reviewed in Hardie, G. and S. Hanks (1995) The Protein Kinase Facts Book, Vol I, pp. 17-20 Academic Press,
30 San Diego CA.). In particular, two protein kinase signature sequences have been identified in the kinase domain, the first containing an active site lysine residue involved in ATP binding, and the second containing an aspartate residue important for catalytic activity. If a protein analyzed includes the two protein kinase signatures, the probability of that protein being a protein kinase is close to 100% (PROSITE: PDOC00100, November 1995).

35 **Protein Tyrosine Kinases**

Protein tyrosine kinases (PTKs) may be classified as either transmembrane, receptor PTKs or nontransmembrane, nonreceptor PTK proteins. Transmembrane tyrosine kinases function as receptors for most growth factors. Growth factors bind to the receptor tyrosine kinase (RTK), which causes the receptor to phosphorylate itself (autophosphorylation) and specific intracellular second messenger proteins. Growth factors (GF) that associate with receptor PTKs include epidermal GF, platelet-derived GF, fibroblast GF, hepatocyte GF, insulin and insulin-like GFs, nerve GF, vascular endothelial GF, and macrophage colony stimulating factor.

Nontransmembrane, nonreceptor PTKs lack transmembrane regions and, instead, form signaling complexes with the cytosolic domains of plasma membrane receptors. Receptors that function through non-receptor PTKs include those for cytokines and hormones (growth hormone and prolactin), and antigen-specific receptors on T and B lymphocytes.

Many PTKs were first identified as oncogene products in cancer cells in which PTK activation was no longer subject to normal cellular controls. In fact, about one third of the known oncogenes encode PTKs. Furthermore, cellular transformation (oncogenesis) is often accompanied by increased tyrosine phosphorylation activity (Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). Regulation of PTK activity may therefore be an important strategy in controlling some types of cancer.

Protein Serine/Threonine Kinases

Protein serine/threonine kinases (STKs) are nontransmembrane proteins. A subclass of STKs are known as ERKs (extracellular signal regulated kinases) or MAPs (mitogen-activated protein kinases) and are activated after cell stimulation by a variety of hormones and growth factors. Cell stimulation induces a signaling cascade leading to phosphorylation of MEK (MAP/ERK kinase) which, in turn, activates ERK via serine and threonine phosphorylation. A varied number of proteins represent the downstream effectors for the active ERK and implicate it in the control of cell proliferation and differentiation, as well as regulation of the cytoskeleton. Activation of ERK is normally transient, and cells possess dual specificity phosphatases that are responsible for its down-regulation. Also, numerous studies have shown that elevated ERK activity is associated with some cancers. Other STKs include the second messenger dependent protein kinases such as the cyclic-AMP dependent protein kinases (PKA), calcium-calmodulin (CaM) dependent protein kinases, and the mitogen-activated protein kinases (MAP); the cyclin-dependent protein kinases; checkpoint and cell cycle kinases; Numb-associated kinase (Nak); human Fused (hFu); proliferation-related kinases; 5'-AMP-activated protein kinases; and kinases involved in apoptosis.

One member of the ERK family of MAP kinases, ERK 7, is a novel 61-kDa protein that has motif similarities to ERK1 and ERK2, but is not activated by extracellular stimuli as are ERK1 and ERK2 nor by the common activators, c-Jun N-terminal kinase (JNK) and p38 kinase. ERK7 regulates

its nuclear localization and inhibition of growth through its C-terminal tail, not through the kinase domain as is typical with other MAP kinases (Abe, M.K. (1999) Mol. Cell. Biol. 19:1301-1312).

The second messenger dependent protein kinases primarily mediate the effects of second messengers such as cyclic AMP (cAMP), cyclic GMP, inositol triphosphate, phosphatidylinositol, 5 3,4,5-triphosphate, cyclic ADP ribose, arachidonic acid, diacylglycerol and calcium-calmodulin. The PKAs are involved in mediating hormone-induced cellular responses and are activated by cAMP produced within the cell in response to hormone stimulation. cAMP is an intracellular mediator of hormone action in all animal cells that have been studied. Hormone-induced cellular responses include thyroid hormone secretion, cortisol secretion, progesterone secretion, glycogen breakdown, 10 bone resorption, and regulation of heart rate and force of heart muscle contraction. PKA is found in all animal cells and is thought to account for the effects of cAMP in most of these cells. Altered PKA expression is implicated in a variety of disorders and diseases including cancer, thyroid disorders, diabetes, atherosclerosis, and cardiovascular disease (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, New York NY, pp. 416-431, 1887).

15 The casein kinase I (CKI) gene family is another subfamily of serine/threonine protein kinases. This continuously expanding group of kinases have been implicated in the regulation of numerous cytoplasmic and nuclear processes, including cell metabolism and DNA replication and repair. CKI enzymes are present in the membranes, nucleus, cytoplasm and cytoskeleton of eukaryotic cells, and on the mitotic spindles of mammalian cells (Fish, K.J. et al. (1995) J. Biol. 20 Chem. 270:14875-14883).

The CKI family members all have a short amino-terminal domain of 9-76 amino acids, a highly conserved kinase domain of 284 amino acids, and a variable carboxyl-terminal domain that ranges from 24 to over 200 amino acids in length (Cegielska, A. et al. (1998) J. Biol. Chem. 273:1357-1364). The CKI family is comprised of highly related proteins, as seen by the identification 25 of isoforms of casein kinase I from a variety of sources. There are at least five mammalian isoforms, α , β , γ , δ , and ϵ . Fish et al. identified CKI-epsilon from a human placenta cDNA library. It is a basic protein of 416 amino acids and is closest to CKI-delta. Through recombinant expression, it was determined to phosphorylate known CKI substrates and was inhibited by the CKI-specific inhibitor CKI-7. The human gene for CKI-epsilon was able to rescue yeast with a slow-growth phenotype 30 caused by deletion of the yeast CKI locus, HRR250 (Fish et al., *supra*).

The mammalian circadian mutation tau was found to be a semidominant autosomal allele of CKI-epsilon that markedly shortens period length of circadian rhythms in Syrian hamsters. The tau locus is encoded by casein kinase I-epsilon, which is also a homolog of the *Drosophila* circadian gene double-time. Studies of both the wildtype and tau mutant CKI-epsilon enzyme indicated that the 35 mutant enzyme has a noticeable reduction in the maximum velocity and autophosphorylation state.

Further, *in vitro*, CKI-epsilon is able to interact with mammalian PERIOD proteins, while the mutant enzyme is deficient in its ability to phosphorylate PERIOD. Lowrey et al. have proposed that CKI-epsilon plays a major role in delaying the negative feedback signal within the transcription-translation-based autoregulatory loop that composes the core of the circadian mechanism. Therefore 5 the CKI-epsilon enzyme is an ideal target for pharmaceutical compounds influencing circadian rhythms, jet-lag and sleep, in addition to other physiologic and metabolic processes under circadian regulation (Lowrey, P.L. et al. (2000) *Science* 288:483-491).

Homeodomain-interacting protein kinases (HIPKs) are serine/threonine kinases and novel members of the DYRK kinase subfamily (Hofmann, T.G. et al. (2000) *Biochimie* 82:1123-1127). 10 HIPKs contain a conserved protein kinase domain separated from a domain that interacts with homeoproteins. HIPKs are nuclear kinases, and HIPK2 is highly expressed in neuronal tissue (Kim, Y.H. et al. (1998) *J. Biol. Chem.* 273:25875-25879; Wang, Y. et al. (2001) *Biochim. Biophys. Acta* 1518:168-172). HIPKs act as corepressors for homeodomain transcription factors. This corepressor activity is seen in posttranslational modifications such as ubiquitination and phosphorylation, each of 15 which are important in the regulation of cellular protein function (Kim, Y.H. et al. (1999) *Proc. Natl. Acad. Sci. USA* 96:12350-12355).

The human h-warts protein, a homolog of *Drosophila* warts tumor suppressor gene, maps to chromosome 6q24-25.1. It has a serine/threonine kinase domain and is localized to centrosomes in interphase cells. It is involved in mitosis and functions as a component of the mitotic apparatus 20 (Nishiyama, Y. et al. (1999) *FEBS Lett.* 459:159-165).

Calcium-Calmodulin Dependent Protein Kinases

Calcium-calmodulin dependent (CaM) kinases are involved in regulation of smooth muscle contraction, glycogen breakdown (phosphorylase kinase), and neurotransmission (CaM kinase I and CaM kinase II). CaM dependent protein kinases are activated by calmodulin, an intracellular calcium receptor, in response to the concentration of free calcium in the cell. Many CaM kinases are also activated by phosphorylation. Some CaM kinases are also activated by autophosphorylation or by other regulatory kinases. CaM kinase I phosphorylates a variety of substrates including the neurotransmitter-related proteins synapsin I and II, the gene transcription regulator, CREB, and the cystic fibrosis conductance regulator protein, CFTR (Haribabu, B. et al. (1995) *EMBO J.* 14:3679-3686). CaM kinase II also phosphorylates synapsin at different sites and controls the synthesis of catecholamines in the brain through phosphorylation and activation of tyrosine hydroxylase. CaM kinase II controls the synthesis of catecholamines and serotonin, through phosphorylation/activation of tyrosine hydroxylase and tryptophan hydroxylase, respectively (Fujisawa, H. (1990) *BioEssays* 12:27-29). The mRNA encoding a calmodulin-binding protein kinase-like protein was found to be 35 enriched in mammalian forebrain. This protein is associated with vesicles in both axons and

dendrites and accumulates largely postnatally. The amino acid sequence of this protein is similar to CaM-dependent STKs, and the protein binds calmodulin in the presence of calcium (Godbout, M. et al. (1994) J. Neurosci. 14:1-13).

Mitogen-Activated Protein Kinases

5 The mitogen-activated protein kinases (MAP), which mediate signal transduction from the cell surface to the nucleus via phosphorylation cascades, are another STK family that regulates intracellular signaling pathways. Several subgroups have been identified, and each manifests different substrate specificities and responds to distinct extracellular stimuli (Egan, S.E. and R.A. Weinberg (1993) Nature 365:781-783). There are three kinase modules comprising the MAP kinase 10 cascade: MAPK (MAP), MAPK kinase (MAP2K, MAPKK, or MKK), and MKK kinase (MAP3K, MAPKKK, OR MEKK) (Wang,X.S. et al (1998) Biochem. Biophys. Res. Commun. 253:33-37). The extracellular-regulated kinase (ERK) pathway is activated by growth factors and mitogens, for example, epidermal growth factor (EGF), ultraviolet light, hyperosmolar medium, heat shock, or endotoxic lipopolysaccharide (LPS). The closely related though distinct parallel pathways, the c-Jun 15 N-terminal kinase (JNK), or stress-activated kinase (SAPK) pathway, and the p38 kinase pathway are activated by stress stimuli and proinflammatory cytokines such as tumor necrosis factor (TNF) and interleukin-1 (IL-1). Altered MAP kinase expression is implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development.. MAP kinase signaling pathways are present in mammalian cells as well as in yeast.

20 Cyclin-Dependent Protein Kinases

The cyclin-dependent protein kinases (CDKs) are STKs that control the progression of cells through the cell cycle. The entry and exit of a cell from mitosis are regulated by the synthesis and destruction of a family of activating proteins called cyclins. Cyclins are small regulatory proteins that bind to and activate CDKs, which then phosphorylate and activate selected proteins involved in the 25 mitotic process. CDKs are unique in that they require multiple inputs to become activated. In addition to cyclin binding, CDK activation requires the phosphorylation of a specific threonine residue and the dephosphorylation of a specific tyrosine residue on the CDK.

Another family of STKs associated with the cell cycle are the NIMA (never in mitosis)-related kinases (Neks). Both CDKs and Neks are involved in duplication, maturation, and separation 30 of the microtubule organizing center, the centrosome, in animal cells (Fry, A.M. et al. (1998) EMBO J. 17:470-481).

Checkpoint and Cell Cycle Kinases

In the process of cell division, the order and timing of cell cycle transitions are under control of cell cycle checkpoints, which ensure that critical events such as DNA replication and chromosome 35 segregation are carried out with precision. If DNA is damaged, e.g. by radiation, a checkpoint

pathway is activated that arrests the cell cycle to provide time for repair. If the damage is extensive, apoptosis is induced. In the absence of such checkpoints, the damaged DNA is inherited by aberrant cells which may cause proliferative disorders such as cancer. Protein kinases play an important role in this process. For example, a specific kinase, checkpoint kinase 1 (Chk1), has been identified in 5 yeast and mammals, and is activated by DNA damage in yeast. Activation of Chk1 leads to the arrest of the cell at the G2/M transition (Sanchez, Y. et al. (1997) Science 277:1497-1501). Specifically, Chk1 phosphorylates the cell division cycle phosphatase CDC25, inhibiting its normal function which is to dephosphorylate and activate the cyclin-dependent kinase Cdc2. Cdc2 activation controls the entry of cells into mitosis (Peng, C.-Y. et al. (1997) Science 277:1501-1505). Thus, activation of 10 Chk1 prevents the damaged cell from entering mitosis. A deficiency in a checkpoint kinase, such as Chk1, may also contribute to cancer by failure to arrest cells with damaged DNA at other checkpoints such as G2/M.

Proliferation-Related Kinases

Proliferation-related kinase is a serum/cytokine inducible STK that is involved in regulation 15 of the cell cycle and cell proliferation in human megakaryocytic cells (Li, B. et al. (1996) J. Biol. Chem. 271:19402-19408). Proliferation-related kinase is related to the polo (derived from *Drosophila* polo gene) family of STKs implicated in cell division. Proliferation-related kinase is downregulated in lung tumor tissue and may be a proto-oncogene whose deregulated expression in normal tissue leads to oncogenic transformation.

5'-AMP-activated protein kinase

A ligand-activated STK protein kinase is 5'-AMP-activated protein kinase (AMPK) (Gao, G. et al. (1996) J. Biol Chem. 271:8675-8681). Mammalian AMPK is a regulator of fatty acid and sterol synthesis through phosphorylation of the enzymes acetyl-CoA carboxylase and hydroxymethylglutaryl-CoA reductase and mediates responses of these pathways to cellular stresses 25 such as heat shock and depletion of glucose and ATP. AMPK is a heterotrimeric complex comprised of a catalytic alpha subunit and two non-catalytic beta and gamma subunits that are believed to regulate the activity of the alpha subunit. Subunits of AMPK have a much wider distribution in non-lipogenic tissues such as brain, heart, spleen, and lung than expected. This distribution suggests that its role may extend beyond regulation of lipid metabolism alone.

Kinases in Apoptosis

Apoptosis is a highly regulated signaling pathway leading to cell death that plays a crucial role in tissue development and homeostasis. Deregulation of this process is associated with the pathogenesis of a number of diseases including autoimmune diseases, neurodegenerative disorders, and cancer. Various STKs play key roles in this process. ZIP kinase is an STK containing a 35 C-terminal leucine zipper domain in addition to its N-terminal protein kinase domain. This

C-terminal domain appears to mediate homodimerization and activation of the kinase as well as interactions with transcription factors such as activating transcription factor, ATF4, a member of the cyclic-AMP responsive element binding protein (ATF/CREB) family of transcriptional factors (Sanjo, H. et al. (1998) J. Biol. Chem. 273:29066-29071). DRAK1 and DRAK2 are STKs that share 5 homology with the death-associated protein kinases (DAP kinases), known to function in interferon- γ induced apoptosis (Sanjo et al., *supra*). Like ZIP kinase, DAP kinases contain a C-terminal protein-protein interaction domain, in the form of ankyrin repeats, in addition to the N-terminal kinase domain. ZIP, DAP, and DRAK kinases induce morphological changes associated with apoptosis when transfected into NIH3T3 cells (Sanjo et al., *supra*). However, deletion of either the 10 N-terminal kinase catalytic domain or the C-terminal domain of these proteins abolishes apoptosis activity, indicating that in addition to the kinase activity, activity in the C-terminal domain is also necessary for apoptosis, possibly as an interacting domain with a regulator or a specific substrate.

RICK is another STK recently identified as mediating a specific apoptotic pathway involving the death receptor, CD95 (Inohara, N. et al. (1998) J. Biol. Chem. 273:12296-12300). CD95 is a 15 member of the tumor necrosis factor receptor superfamily and plays a critical role in the regulation and homeostasis of the immune system (Nagata, S. (1997) Cell 88:355-365). The CD95 receptor signaling pathway involves recruitment of various intracellular molecules to a receptor complex following ligand binding. This process includes recruitment of the cysteine protease caspase-8 which, in turn, activates a caspase cascade leading to cell death. RICK is composed of an N-terminal 20 kinase catalytic domain and a C-terminal "caspase-recruitment" domain that interacts with caspase-like domains, indicating that RICK plays a role in the recruitment of caspase-8. This interpretation is supported by the fact that the expression of RICK in human 293T cells promotes activation of caspase-8 and potentiates the induction of apoptosis by various proteins involved in the CD95 apoptosis pathway (Inohara et al., *supra*).

25 **Mitochondrial Protein Kinases**

A novel class of eukaryotic kinases, related by sequence to prokaryotic histidine protein kinases, are the mitochondrial protein kinases (MPKs) which seem to have no sequence similarity with other eukaryotic protein kinases. These protein kinases are located exclusively in the mitochondrial matrix space and may have evolved from genes originally present in respiration-dependent bacteria which were endocytosed by primitive eukaryotic cells. MPKs are responsible for phosphorylation and inactivation of the branched-chain alpha-ketoacid dehydrogenase and pyruvate dehydrogenase complexes (Harris, R.A. et al. (1995) Adv. Enzyme Regul. 34:147-162). Five MPKs have been identified. Four members correspond to pyruvate dehydrogenase kinase isozymes, regulating the activity of the pyruvate dehydrogenase complex, which is an important regulatory 30 enzyme at the interface between glycolysis and the citric acid cycle. The fifth member corresponds to 35

a branched-chain alpha-ketoacid dehydrogenase kinase, important in the regulation of the pathway for the disposal of branched-chain amino acids. (Harris, R.A. et al. (1997) *Adv. Enzyme Regul.* 37:271-293). Both starvation and the diabetic state are known to result in a great increase in the activity of the pyruvate dehydrogenase kinase in the liver, heart and muscle of the rat. This increase contributes 5 in both disease states to the phosphorylation and inactivation of the pyruvate dehydrogenase complex and conservation of pyruvate and lactate for gluconeogenesis (Harris (1995) *supra*).

KINASES WITH NON-PROTEIN SUBSTRATES

Lipid and Inositol kinases

Lipid kinases phosphorylate hydroxyl residues on lipid head groups. A family of kinases 10 involved in phosphorylation of phosphatidylinositol (PI) has been described, each member phosphorylating a specific carbon on the inositol ring (Leevers, S.J. et al. (1999) *Curr. Opin. Cell. Biol.* 11:219-225). The phosphorylation of phosphatidylinositol is involved in activation of the protein kinase C signaling pathway. The inositol phospholipids (phosphoinositides) intracellular 15 signaling pathway begins with binding of a signaling molecule to a G-protein linked receptor in the plasma membrane. This leads to the phosphorylation of phosphatidylinositol (PI) residues on the inner side of the plasma membrane by inositol kinases, thus converting PI residues to the biphosphate state (PIP₂). PIP₂ is then cleaved into inositol triphosphate (IP₃) and diacylglycerol. These two 20 products act as mediators for separate signaling pathways. Cellular responses that are mediated by these pathways are glycogen breakdown in the liver in response to vasopressin, smooth muscle contraction in response to acetylcholine, and thrombin-induced platelet aggregation.

PI 3-kinase (PI3K), which phosphorylates the D3 position of PI and its derivatives, has a central role in growth factor signal cascades involved in cell growth, differentiation, and metabolism. PI3K is a heterodimer consisting of an adapter subunit and a catalytic subunit. The adapter subunit acts as a scaffolding protein, interacting with specific tyrosine-phosphorylated proteins, lipid 25 moieties, and other cytosolic factors. When the adapter subunit binds tyrosine phosphorylated targets, such as the insulin responsive substrate (IRS)-1, the catalytic subunit is activated and converts PI (4,5) bisphosphate (PIP₂) to PI (3,4,5) P₃ (PIP₃). PIP₃ then activates a number of other proteins, including PKA, protein kinase B (PKB), protein kinase C (PKC), glycogen synthase kinase (GSK)-3, and p70 ribosomal s6 kinase. PI3K also interacts directly with the cytoskeletal organizing proteins, 30 Rac, rho, and cdc42 (Shepherd, P.R. et al. (1998) *Biochem. J.* 333:471-490). Animal models for diabetes, such as *obese* and *fat* mice, have altered PI3K adapter subunit levels. Specific mutations in the adapter subunit have also been found in an insulin-resistant Danish population, suggesting a role for PI3K in type-2 diabetes (Shepard, *supra*).

An example of lipid kinase phosphorylation activity is the phosphorylation of 35 D-erythro-sphingosine to the sphingolipid metabolite, sphingosine-1-phosphate (SPP). SPP has

emerged as a novel lipid second-messenger with both extracellular and intracellular actions (Kohama, T. et al. (1998) J. Biol. Chem. 273:23722-23728). Extracellularly, SPP is a ligand for the G-protein coupled receptor EDG-1 (endothelial-derived, G-protein coupled receptor). Intracellularly, SPP regulates cell growth, survival, motility, and cytoskeletal changes. SPP levels are regulated by sphingosine kinases that specifically phosphorylate D-erythro-sphingosine to SPP. The importance of sphingosine kinase in cell signaling is indicated by the fact that various stimuli, including platelet-derived growth factor (PDGF), nerve growth factor, and activation of protein kinase C, increase cellular levels of SPP by activation of sphingosine kinase, and the fact that competitive inhibitors of the enzyme selectively inhibit cell proliferation induced by PDGF (Kohama et al., 10 *supra*).

Purine Nucleotide Kinases

The purine nucleotide kinases, adenylate kinase (ATP:AMP phosphotransferase, or AdK) and guanylate kinase (ATP:GMP phosphotransferase, or GuK) play a key role in nucleotide metabolism and are crucial to the synthesis and regulation of cellular levels of ATP and GTP, respectively. These two molecules are precursors in DNA and RNA synthesis in growing cells and provide the primary source of biochemical energy in cells (ATP), and signal transduction pathways (GTP). Inhibition of various steps in the synthesis of these two molecules has been the basis of many antiproliferative drugs for cancer and antiviral therapy (Pillwein, K. et al. (1990) Cancer Res. 50:1576-1579).

AdK is found in almost all cell types and is especially abundant in cells having high rates of ATP synthesis and utilization such as skeletal muscle. In these cells AdK is physically associated with mitochondria and myofibrils, the subcellular structures that are involved in energy production and utilization, respectively. Recent studies have demonstrated a major function for AdK in transferring high energy phosphoryls from metabolic processes generating ATP to cellular components consuming ATP (Zeleznikar, R.J. et al. (1995) J. Biol. Chem. 270:7311-7319). Thus AdK may have a pivotal role in maintaining energy production in cells, particularly those having a high rate of growth or metabolism such as cancer cells, and may provide a target for suppression of its activity in order to treat certain cancers. Alternatively, reduced AdK activity may be a source of various metabolic, muscle-energy disorders that can result in cardiac or respiratory failure and may be treatable by increasing AdK activity.

GuK, in addition to providing a key step in the synthesis of GTP for RNA and DNA synthesis, also fulfills an essential function in signal transduction pathways of cells through the regulation of GDP and GTP. Specifically, GTP binding to membrane associated G proteins mediates the activation of cell receptors, subsequent intracellular activation of adenyl cyclase, and production of the second messenger, cyclic AMP. GDP binding to G proteins inhibits these processes. GDP and GTP levels also control the activity of certain oncogenic proteins such as p21^{ras} known to be involved

in control of cell proliferation and oncogenesis (Bos, J.L. (1989) *Cancer Res.* 49:4682-4689). High ratios of GTP:GDP caused by suppression of GuK cause activation of p21^{ras} and promote oncogenesis. Increasing GuK activity to increase levels of GDP and reduce the GTP:GDP ratio may provide a therapeutic strategy to reverse oncogenesis.

5 GuK is an important enzyme in the phosphorylation and activation of certain antiviral drugs useful in the treatment of herpes virus infections. These drugs include the guanine homologs acyclovir and buclovir (Miller, W.H. and R.L. Miller (1980) *J. Biol. Chem.* 255:7204-7207; Stenberg, K. et al. (1986) *J. Biol. Chem.* 261:2134-2139). Increasing GuK activity in infected cells may provide a therapeutic strategy for augmenting the effectiveness of these drugs and possibly for
10 reducing the necessary dosages of the drugs.

Pyrimidine Kinases

The pyrimidine kinases are deoxycytidine kinase and thymidine kinase 1 and 2.

Deoxycytidine kinase is located in the nucleus, and thymidine kinase 1 and 2 are found in the cytosol (Johansson, M. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:11941-11945). Phosphorylation of
15 deoxyribonucleosides by pyrimidine kinases provides an alternative pathway for *de novo* synthesis of DNA precursors. The role of pyrimidine kinases, like purine kinases, in phosphorylation is critical to the activation of several chemotherapeutically important nucleoside analogues (Arner E.S. and S. Eriksson (1995) *Pharmacol. Ther.* 67:155-186).

PHOSPHATASES

20 Protein phosphatases are generally characterized as either serine/threonine- or tyrosine-specific based on their preferred phospho-amino acid substrate. However, some phosphatases (DSPs, for dual specificity phosphatases) can act on phosphorylated tyrosine, serine, or threonine residues. The protein serine/threonine phosphatases (PSPs) are important regulators of many cAMP-mediated hormone responses in cells. Protein tyrosine phosphatases (PTPs) play a significant role in cell cycle
25 and cell signaling processes. Another family of phosphatases is the acid phosphatase or histidine acid phosphatase (HAP) family whose members hydrolyze phosphate esters at acidic pH conditions.

PSPs are found in the cytosol, nucleus, and mitochondria and in association with cytoskeletal and membranous structures in most tissues, especially the brain. Some PSPs require divalent cations, such as Ca²⁺ or Mn²⁺, for activity. PSPs play important roles in glycogen metabolism, muscle
30 contraction, protein synthesis, T cell function, neuronal activity, oocyte maturation, and hepatic metabolism (reviewed in Cohen, P. (1989) *Annu. Rev. Biochem.* 58:453-508). PSPs can be separated into two classes. The PPP class includes PP1, PP2A, PP2B/calcineurin, PP4, PP5, PP6, and PP7. Members of this class are composed of a homologous catalytic subunit bearing a very highly
conserved signature sequence, coupled with one or more regulatory subunits (PROSITE
35 PDOC00115). Further interactions with scaffold and anchoring molecules determine the intracellular

localization of PSPs and substrate specificity. The PPM class consists of several closely related isoforms of PP2C and is evolutionarily unrelated to the PPP class.

PP1 dephosphorylates many of the proteins phosphorylated by cyclic AMP-dependent protein kinase (PKA) and is an important regulator of many cAMP-mediated hormone responses in cells. A number of isoforms have been identified, with the alpha and beta forms being produced by alternative splicing of the same gene. Both ubiquitous and tissue-specific targeting proteins for PP1 have been identified. In the brain, inhibition of PP1 activity by the dopamine and adenosine 3',5'-monophosphate-regulated phosphoprotein of 32kDa (DARPP-32) is necessary for normal dopamine response in neostriatal neurons (reviewed in Price, N.E. and M.C. Mumby (1999) *Curr. Opin. Neurobiol.* 9:336-342). PP1, along with PP2A, has been shown to limit motility in microvascular endothelial cells, suggesting a role for PSPs in the inhibition of angiogenesis (Gabel, S. et al. (1999) *Otolaryngol. Head Neck Surg.* 121:463-468).

PP2A is the main serine/threonine phosphatase. The core PP2A enzyme consists of a single 36 kDa catalytic subunit (C) associated with a 65 kDa scaffold subunit (A), whose role is to recruit additional regulatory subunits (B). Three gene families encoding B subunits are known (PR55, PR61, and PR72), each of which contain multiple isoforms, and additional families may exist (Millward, T.A. et al. (1999) *Trends Biosci.* 24:186-191). These "B-type" subunits are cell type- and tissue-specific and determine the substrate specificity, enzymatic activity, and subcellular localization of the holoenzyme. The PR55 family is highly conserved and bears a conserved motif (PROSITE PDOC00785). PR55 increases PP2A activity toward mitogen-activated protein kinase (MAPK) and MAPK kinase (MEK). PP2A dephosphorylates the MAPK active site, inhibiting the cell's entry into mitosis. Several proteins can compete with PR55 for PP2A core enzyme binding, including the CKII kinase catalytic subunit, polyomavirus middle and small T antigens, and SV40 small t antigen. Viruses may use this mechanism to commandeer PP2A and stimulate progression of the cell through the cell cycle (Pallas, D.C. et al. (1992) *J. Virol.* 66:886-893). Altered MAP kinase expression is also implicated in a variety of disease conditions including cancer, inflammation, immune disorders, and disorders affecting growth and development. PP2A, in fact, can dephosphorylate and modulate the activities of more than 30 protein kinases *in vitro*, and other evidence suggests that the same is true *in vivo* for such kinases as PKB, PKC, the calmodulin-dependent kinases, ERK family MAP kinases, cyclin-dependent kinases, and the I_KB kinases (reviewed in Millward et al., *supra*). PP2A is itself a substrate for CKI and CKII kinases, and can be stimulated by polycationic macromolecules. A PP2A-like phosphatase is necessary to maintain the G1 phase destruction of mammalian cyclins A and B (Bastians, H. et al. (1999) *Mol. Biol. Cell* 10:3927-3941). PP2A is a major activity in the brain and is implicated in regulating neurofilament stability and normal neural function, particularly the phosphorylation of the microtubule-associated protein tau. Hyperphosphorylation of tau has been

proposed to lead to the neuronal degeneration seen in Alzheimer's disease (reviewed in Price and Mumby, *supra*).

PP2B, or calcineurin, is a Ca^{2+} -activated dimeric phosphatase and is particularly abundant in the brain. It consists of catalytic and regulatory subunits, and is activated by the binding of the 5 calcium/calmodulin complex. Calcineurin is the target of the immunosuppressant drugs cyclosporine and FK506. Along with other cellular factors, these drugs interact with calcineurin and inhibit phosphatase activity. In T cells, this blocks the calcium dependent activation of the NF-AT family of transcription factors, leading to immunosuppression. This family is widely distributed, and it is likely that calcineurin regulates gene expression in other tissues as well. In neurons, calcineurin modulates 10 functions which range from the inhibition of neurotransmitter release to desensitization of postsynaptic NMDA-receptor coupled calcium channels to long term memory (reviewed in Price and Mumby, *supra*).

Other members of the PPP class have recently been identified (Cohen, P.T. (1997) Trends Biochem. Sci. 22:245-251). One of them, PP5, contains regulatory domains with tetratricopeptide 15 repeats. It can be activated by polyunsaturated fatty acids and anionic phospholipids *in vitro* and appears to be involved in a number of signaling pathways, including those controlled by atrial natriuretic peptide or steroid hormones (reviewed in Andreeva, A.V. and M.A. Kutuzov (1999) Cell Signal. 11:555-562).

PP2C is a ~42kDa monomer with broad substrate specificity and is dependent on divalent 20 cations (mainly Mn^{2+} or Mg^{2+}) for its activity. PP2C proteins share a conserved N-terminal region with an invariant DGH motif, which contains an aspartate residue involved in cation binding (PROSITE PDOC00792). Targeting proteins and mechanisms regulating PP2C activity have not been identified. PP2C has been shown to inhibit the stress-responsive p38 and Jun kinase (JNK) pathways (Takekawa, M. et al. (1998) EMBO J. 17:4744-4752).

In contrast to PSPs, tyrosine-specific phosphatases (PTPs) are generally monomeric proteins 25 of very diverse size (from 20kDa to greater than 100kDa) and structure that function primarily in the transduction of signals across the plasma membrane. PTPs are categorized as either soluble phosphatases or transmembrane receptor proteins that contain a phosphatase domain. All PTPs share a conserved catalytic domain of about 300 amino acids which contains the active site. The active site 30 consensus sequence includes a cysteine residue which executes a nucleophilic attack on the phosphate moiety during catalysis (Neel, B.G. and N.K. Tonks (1997) Curr. Opin. Cell Biol. 9:193-204). Receptor PTPs are made up of an N-terminal extracellular domain of variable length, a 35 transmembrane region, and a cytoplasmic region that generally contains two copies of the catalytic domain. Although only the first copy seems to have enzymatic activity, the second copy apparently affects the substrate specificity of the first. The extracellular domains of some receptor PTPs contain

fibronectin-like repeats, immunoglobulin-like domains, MAM domains (an extracellular motif likely to have an adhesive function), or carbonic anhydrase-like domains (PROSITE PDOC 00323). This wide variety of structural motifs accounts for the diversity in size and specificity of PTPs.

PTPs play important roles in biological processes such as cell adhesion, lymphocyte

5 activation, and cell proliferation. PTPs μ and κ are involved in cell-cell contacts, perhaps regulating cadherin/catenin function. A number of PTPs affect cell spreading, focal adhesions, and cell motility, most of them via the integrin/tyrosine kinase signaling pathway (reviewed in Neel and Tonks, *supra*). CD45 phosphatases regulate signal transduction and lymphocyte activation (Ledbetter, J.A. et al. (1988) Proc. Natl. Acad. Sci. USA 85:8628-8632). Soluble PTPs containing Src-homology-2
10 domains have been identified (SHPs), suggesting that these molecules might interact with receptor tyrosine kinases. SHP-1 regulates cytokine receptor signaling by controlling the Janus family PTKs in hematopoietic cells, as well as signaling by the T-cell receptor and c-Kit (reviewed in Neel and Tonks, *supra*). M-phase inducer phosphatase plays a key role in the induction of mitosis by dephosphorylating and activating the PTK CDC2, leading to cell division (Sadhu, K. et al. (1990)
15 Proc. Natl. Acad. Sci. USA 87:5139-5143). In addition, the genes encoding at least eight PTPs have been mapped to chromosomal regions that are translocated or rearranged in various neoplastic conditions, including lymphoma, small cell lung carcinoma, leukemia, adenocarcinoma, and neuroblastoma (reviewed in Charbonneau, H. and N.K. Tonks (1992) Annu. Rev. Cell Biol. 8:463-493). The PTP enzyme active site comprises the consensus sequence of the MTM1 gene family. The
20 MTM1 gene is responsible for X-linked recessive myotubular myopathy, a congenital muscle disorder that has been linked to Xq28 (Kioschis, P. et al., (1998) Genomics 54:256-266). Many PTKs are encoded by oncogenes, and it is well known that oncogenesis is often accompanied by increased tyrosine phosphorylation activity. It is therefore possible that PTPs may serve to prevent or reverse cell transformation and the growth of various cancers by controlling the levels of tyrosine
25 phosphorylation in cells. This is supported by studies showing that overexpression of PTP can suppress transformation in cells and that specific inhibition of PTP can enhance cell transformation (Charbonneau and Tonks, *supra*).

Dual specificity phosphatases (DSPs) are structurally more similar to the PTPs than the PSPs.

DSPs bear an extended PTP active site motif with an additional 7 amino acid residues. DSPs are
30 primarily associated with cell proliferation and include the cell cycle regulators cdc25A, B, and C. The phosphatases DUSP1 and DUSP2 inactivate the MAPK family members ERK (extracellular signal-regulated kinase), JNK (c-Jun N-terminal kinase), and p38 on both tyrosine and threonine residues (PROSITE PDOC 00323, *supra*). In the activated state, these kinases have been implicated in neuronal differentiation, proliferation, oncogenic transformation, platelet aggregation, and
35 apoptosis. Thus, DSPs are necessary for proper regulation of these processes (Muda, M. et al. (1996)

J. Biol. Chem. 271:27205-27208). The tumor suppressor PTEN is a DSP that also shows lipid phosphatase activity. It seems to negatively regulate interactions with the extracellular matrix and maintains sensitivity to apoptosis. PTEN has been implicated in the prevention of angiogenesis (Giri, D. and M. Ittmann (1999) Hum. Pathol. 30:419-424) and abnormalities in its expression are 5 associated with numerous cancers (reviewed in Tamura, M. et al. (1999) J. Natl. Cancer Inst. 91:1820-1828).

Histidine acid phosphatase (HAP; EXPASY EC 3.1.3.2), also known as acid phosphatase, hydrolyzes a wide spectrum of substrates including alkyl, aryl, and acyl orthophosphate monoesters and phosphorylated proteins at low pH. HAPs share two regions of conserved sequences, each 10 centered around a histidine residue which is involved in catalytic activity. Members of the HAP family include lysosomal acid phosphatase (LAP) and prostatic acid phosphatase (PAP), both sensitive to inhibition by L-tartrate (PROSITE PDOC00538).

Synaptojanin, a polyphosphoinositide phosphatase, dephosphorylates phosphoinositides at positions 3, 4 and 5 of the inositol ring. Synaptojanin is a major presynaptic protein found at clathrin-coated endocytic intermediates in nerve terminals, and binds the clathrin coat-associated protein, EPS15. This binding is mediated by the C-terminal region of synaptojanin-170, which has 3 Asp-Pro-Phe amino acid repeats. Further, this 3 residue repeat had been found to be the binding site for the EH domains of EPS15 (Haffner, C. et al. (1997) FEBS Lett. 419:175-180). Additionally, 15 synaptojanin may potentially regulate interactions of endocytic proteins with the plasma membrane, and be involved in synaptic vesicle recycling (Brodin, L. et al. (2000) Curr. Opin. Neurobiol. 10:312-320). Studies in mice with a targeted disruption in the synaptojanin 1 gene (Synj1) were shown to support coat formation of endocytic vesicles more effectively than was seen in wild-type mice, suggesting that Synj1 can act as a negative regulator of membrane-coat protein interactions. These 20 findings provide genetic evidence for a crucial role of phosphoinositide metabolism in synaptic vesicle recycling (Cremona, O. et al. (1999) Cell 99:179-188).

Expression profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in 30 a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single 35 gene is examined, arrays are employed to detect the expression of a specific gene or its variants.

When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

5 Lung Cancer

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as 10 small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

15 Ovarian Cancer

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rate for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

20 Breast Cancer

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (K. Gish (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in 25 which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou CM et al. (2000) Nature 406:747-752).

30 Breast cancer is a genetic disease commonly caused by mutations in cellular disease. Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, *supra*). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to noninherited mutations that occur in breast epithelial cells.

35 A good deal is already known about the expression of specific genes associated with breast

cancer. For example, the relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie et al., *supra*, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in 5 breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been 10 implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S.S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled 15 protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed in human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou Z et al. (1998) Int. J. Cancer 78:95-99; Chen, L et al. (1990) Oncogene 5:1391-1395; Ulrix W et al. (1999) FEBS Lett 20 455:23-26; Sager, R et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, SW et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Estrogen stimulation plays a critical role in the development of normal mammary epithelium. Estradiol has a direct mitogenic effect on breast cancer cells, causing them to divide more rapidly by shortening their cell cycle. Also, estradiol induces a large number of enzymes and other proteins 25 involved in nucleic acid synthesis in isolated breast cancer cell lines. Estradiol may increase the expression of the EGF receptor in response to TGF- α and EGF. In addition, estrogens may promote proliferation of tumor cells by inducing the synthesis of TGF- α and EGF, and may block growth factors that would normally inhibit tumor cell growth. Estrogen receptor (ER) has been investigated extensively as a prognostic marker in breast cancer. Patients whose tumors display high levels of 30 estrogen receptor have a significantly better prognosis than patients with receptor-negative tumors. When ER is lost or cells expressing the ER are selected against by therapeutic treatments, the tumor becomes more aggressive.

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it 35 has been shown that these cell lines retain many of the properties of their parental tumors for lengthy

culture periods (Wistuba II et al. (1998) Clin Cancer Res 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Tumor cells stimulate the formation of stroma that secretes various mediators, such as growth factors, cytokines, and proteases, which are critical for tumor growth. For instance, serum tumor necrosis factor alpha (TNF- α) increased in the circulation of patients with malignancy. Thus it is likely that the serum TNF- α concentration can serve as a biomarker for staging the invasiveness of breast cancer. Clinically, TNF- α treatment in combination with Interferon-gamma (IFN- γ) may provide a successful approach to overcome the cellular heterogeneity of advanced breast tumors.

TNF- α , also called cachectin, is produced by neutrophils, activated lymphocytes, macrophages, NK cells, LAK cells, astrocytes, endothelial cells, smooth muscle cells, and some transformed cells. TNF- α occurs as a secreted, soluble form and as a membrane-anchored form, both of which are biologically active. Two types of receptors for TNF- α have been described and virtually all cell types studied show the presence of one or both of these receptor types. TNF- α and TNF- β are extremely pleiotropic factors due to the ubiquity of their receptors, to their ability to activate multiple signal transduction path-ways and to their ability to induce or suppress the expression of a wide number of genes. TNF- α and TNF- β play a critical role in mediation of the inflammatory response and in mediation of resistance to infections and tumor growth.

Atherosclerosis

Blood vessel walls are composed of two tissue layers: an endothelial cell (EC) layer which comprises the luminal surface of the vessel, and an underlying vascular smooth muscle cell (VSMC) layer. Through dynamic interactions with each other and with surrounding tissues, the vascular endothelium and smooth muscle tissues maintain vascular tone, control selective permeability of the vascular wall, direct vessel remodeling and angiogenesis, and modulate inflammatory and immune responses.

The inflammatory response is a complex vascular reaction mediated by numerous cytokines, chemokines, growth factors, and other signaling molecules expressed by activated ECs, VSMCs and leukocytes. Inflammation protects the organism during trauma and infection, but can also lead to pathological conditions such as atherosclerosis. The pro-inflammatory cytokines, interleukin (IL)-1 and tumor necrosis factor (TNF), are secreted by a small number of activated macrophages or other cells and can set off a cascade of vascular changes, largely through their ability to alter gene expression patterns in ECs and VSMCs. These vascular changes include vasodilation and increased permeability of microvasculature, edema, and leukocyte extravasation and transmigration across the vessel wall. Ultimately, leukocytes, particularly neutrophils and monocytes/macrophages, accumulate in the extravascular space, where they remove injurious agents by phagocytosis and

oxidative killing, a process accompanied by release of toxic factors, such as proteases and reactive oxygen species.

IL-1 and TNF induce pro-inflammatory, thrombotic, and anti-apoptotic changes in gene expression by signaling through receptors on the surface of ECs and VSMCs; these receptors activate transcription factors such as NF κ B as well as AP-1, IRF-1, and NF-GMa, leading to alterations in gene expression. Genes known to be differentially regulated in EC by IL-1 and TNF include E selectin, VCAM-1, ICAM-1, PAF, I κ B α , IAP-1, MCP-1, eotaxin, ENA-78, G-CSF, A20, ICE, and complement C3 component. A key event in inflammation, adhesion and transmigration of blood leukocytes across the vascular endothelium, for example, is mediated by increased expression of E selectin, P selectin, ICAM-1, and VCAM-1 on activated endothelium.

Human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a human aorta. HAECs have been used as an experimental model for investigating the role of the endothelium in human vascular biology *in vitro*. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation coagulation and thrombosis, atherosclerosis, and inflammation.

Atherosclerosis is a pathological condition characterized by a chronic local inflammatory response within the vessel wall of major arteries. Disease progression results in the formation of atherosclerotic lesions, unstable plaques which occasionally rupture, precipitating a catastrophic thrombotic occlusion of the vessel lumen. Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common causes of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and identifies all potential therapeutic targets for this complex disease has not been achieved. Molecular characterization of atherosclerosis requires identification of the genes that contribute to lesion growth, stability, dissolution, rupture and induction of occlusive vessel thrombi.

Human umbilical vein endothelial cells (HUVEC) also are used in the study of vascular and other physiology. Atherosclerosis is a pathological condition characterized by a chronic local inflammatory response within the vessel wall of major arteries. Disease progression results in the formation of atherosclerotic lesions, unstable plaques which occasionally rupture, precipitating a catastrophic thrombotic occlusion of the vessel lumen. Atherosclerosis and the associated coronary artery disease and cerebral stroke represent the most common causes of death in industrialized nations. Although certain key risk factors have been identified, a full molecular characterization that elucidates the causes and identifies all potential therapeutic targets for this complex disease has not been achieved. Molecular characterization of atherosclerosis requires identification of the genes that contribute to lesion growth, stability, dissolution, rupture and induction of occlusive vessel thrombi.

Tumor necrosis factor-alpha (TNF- α) mediates immune regulation and inflammatory

responses through various intermediates, including protein kinases, protein phosphatases, reactive oxygen intermediates, phospholipases, proteases, sphingomyelinases and transcription factors. TNF- α -related cytokines generate cellular responses including differentiation, proliferation, cell death, and activation of nuclear factor- κ B (NF- κ B) (Smith, C.A. et al. (1994) Cell 76:959-962), through their interaction with distinct cell surface receptors (TNFRs). NF- κ B is a transcription factor that induces genes involved in physiological processes such as response to injury and infection. (For a review of TNF- α in the NF- κ B activation pathway see Bowie and O'Neill (2000) Biochem. Pharmacol. 59:13-23.) IL-1 β and TNF- α induce pro-inflammatory, thrombotic, and anti-apoptotic changes in gene expression by signaling through receptors on the surface of ECs and VSMCs; these receptors activate several transcription factors, leading to alterations in gene expression. Genes known to be differentially regulated in EC by IL-1 β and TNF- α include E-selectin, VCAM-1, ICAM-1, PAF, I κ B α , IAP-1, MCP-1, eotaxin, ENA-78, G-CSF, A20, ICE, and complement C3 component. A key event in inflammation, adhesion and transmigration of blood leukocytes across the vascular endothelium, for example, is mediated by increased expression of E-selectin, P-selectin, ICAM-1, and VCAM-1 on activated endothelium.

TNF- α is upregulated when the endothelium is physically disrupted or functionally perturbed by events such as postischemic reperfusion, acute and chronic inflammation, atherosclerosis, diabetes and chronic arterial hypertension. Inflammatory stimulation sets the stage for later tissue repair. Elevated TNF- α initially increases, and then inhibits, the activity of a number of key enzymes including protein-tyrosine kinase (PTKase) and protein-tyrosine phosphatase (Holden, R.J. et al. (1999) Med. Hypotheses 52:319-23).

Development of atherosclerosis involves inflammatory responses induced by circulating lipoprotein. Lipoproteins, such as low-density lipoprotein (LDL), accumulate in the extracellular space of the vascular intima and undergo modifications including oxidation of LDL to Ox-LDL, most avidly in the sub-endothelial space where circulating antioxidant defenses are less effective. Mononuclear phagocytes enter the intima, differentiate into macrophages, and ingest modified lipids including Ox-LDL. During Ox-LDL uptake, macrophages produce cytokines including TNF- α , as well as interleukin-1 and growth factors (e.g. M-CSF, VEGF, and PDGF-BB), that elicit further events in atherogenesis such as smooth muscle cell proliferation and production of extracellular matrix by vascular endothelium. These macrophages may also activate genes in endothelium and smooth muscle tissue involved in inflammation and tissue differentiation, including superoxide dismutatse (SOD), IL-8, and ICAM-1.

Non-atherosclerotic vascular endothelium not only mediates vascular dilatation but prevents platelet adhesion and activation, blocks thrombin formation, mitigates fibrin deposition, and attenuates adhesion and transmigration of inflammatory leukocytes. When the endothelium is

physically disrupted, or perturbed by events such as postischemic reperfusion, acute and chronic inflammation, atherosclerosis, diabetes and chronic arterial hypertension, it acts in the opposite manner. The perturbed or proinflammatory state is characterised by vaso-constriction, platelet and leukocyte activation and adhesion (involving externalisation, expression and upregulation of, for example, von Willebrand factor, platelet activating factor, P-selectin, ICAM-1, IL-8, MCP-1, and TNF- α), promotion of thrombin formation, coagulation and deposition of fibrin at the vascular wall (expression of tissue factor, PAI-1, and phosphatidyl serine) and, in platelet-leukocyte coaggregates, additional inflammatory interactions via attachment of platelet CD40-ligand to endothelial, monocyte and B-cell CD40. Thrombin formation and inflammatory stimulation set the stage for later tissue repair, but limiting procoagulatory, prothrombotic actions of a dysfunctional vascular endothelium may be the goal of clinical interventions (for review, see Becker et al. (2000) Z Kardiol 89:160-167).

Several investigators have examined changes in vascular cell gene expression associated with various inflammatory diseases or model systems. Examining human umbilical vein endothelial cells (HUVEC) activated by recombinant TNF- α or conditioned medium from activated human primary monocytes, Horrevoets et al. (1999; Blood 93:3418-3431) identified 106 differentially regulated genes. In a similar approach, deVries et al. (2000; J. Biol. Chem. 275:23939-23947) identified 40 differentially regulated genes in umbilical cord artery-derived smooth muscle cells activated by conditioned media from cultured macrophages after stimulation with oxidized LDL particles. In both studies, many of the identified genes were already known to be involved in inflammation. Comparing expression profiles from inflammatory diseased tissues, cultured macrophages, chondrocyte cell lines, primary chondrocytes, and synoviocytes, Heller et al. (1997; Proc. Natl. Acad. Sci. USA 94:2150-2155) identified candidate genes involved in inflammatory responses, including TNF, IL-1 IL-6, IL-8 G-CSF, RANTES, and V-CAM. From this candidate gene set, tissue inhibitor of metalloproteinase 1, ferritin light chain, and manganese superoxide dismutase were found to be differentially expressed in rheumatoid arthritis (RA) relative to inflammatory bowel disease (IBD). Further, IL-3, chemokine Gro α , and metalloproteinase matrix metallo-elastase were expressed in both RA and IBD. Most recently, in an analysis of cultured aortic smooth muscle cells treated with TNF α , Haley et al. (2000; Circulation 102:2185-2189) found a 20-fold increase in eotaxin, an eosinophil chemotactic factor. The overexpression of eotaxin and its receptor CCR3 in atherosclerotic lesions was confirmed by northern analysis.

Colon Cancer

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined

sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations 5 in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered 10 sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Tangier Disease

15 Tangier disease (TD) is a genetic disorder characterized by the near absence of circulating high density lipoprotein (HDL) and the accumulation of cholesterol esters in many tissues, including tonsils, lymph nodes, liver, spleen, thymus, and intestine. Low levels of HDL represent a clear predictor of premature coronary artery disease and homozygous TD correlates with a four- to six-fold increase in cardiovascular disease compared to controls. HDL plays a cardio-protective role in 20 reverse cholesterol transport, the flux of cholesterol from peripheral cells such as tissue macrophages through plasma lipoproteins to the liver. The HDL protein, apolipoprotein A-I, plays a major role in this process, interacting with the cell surface to remove excess cholesterol and phospholipids. This pathway is severely impaired in TD and the defect lies in a specific gene, the ABC1 transporter. This gene is a member of the family of ATP-binding cassette transporters, which utilize ATP hydrolysis to 25 transport a variety of substrates across membranes.

Steroids

30 Steroids are a class of lipid-soluble molecules, including cholesterol, bile acids, vitamin D, and hormones, that share a common four-ring structure based on cyclopentanoperhydrophenanthrene and that carry out a wide variety of functions. Cholesterol, for example, is a component of cell membranes that controls membrane fluidity. It is also a precursor for bile acids which solubilize lipids and facilitate absorption in the small intestine during digestion. Vitamin D regulates the absorption of calcium in the small intestine and controls the concentration of calcium in plasma. Steroid hormones, produced by the adrenal cortex, ovaries, and testes, include glucocorticoids, 35 mineralocorticoids, androgens, and estrogens. They control various biological processes by binding to intracellular receptors that regulate transcription of specific genes in the nucleus.

Glucocorticoids, for example, increase blood glucose concentrations by regulation of gluconeogenesis in the liver, increase blood concentrations of fatty acids by promoting lipolysis in adipose tissues, modulate sensitivity to catecholamines in the central nervous system, and reduce inflammation. The principal mineralocorticoid, aldosterone, is produced by the adrenal cortex and acts on cells of the distal tubules of the kidney to enhance sodium ion reabsorption. Androgens, produced by the interstitial cells of Leydig in the testis, include the male sex hormone testosterone, which triggers changes at puberty, the production of sperm and maintenance of secondary sexual characteristics. Female sex hormones, estrogen and progesterone, are produced by the ovaries and also by the placenta and adrenal cortex of the fetus during pregnancy. Estrogen regulates female reproductive processes and secondary sexual characteristics. Progesterone regulates changes in the endometrium during the menstrual cycle and pregnancy.

Steroid hormones are widely used for fertility control and in anti-inflammatory treatments for physical injuries and diseases such as arthritis, asthma, and autoimmune disorders. Progesterone, a naturally occurring progestin, is primarily used to treat amenorrhea, abnormal uterine bleeding, or as a contraceptive. Endogenous progesterone is responsible for inducing secretory activity in the endometrium of the estrogen-primed uterus in preparation for the implantation of a fertilized egg and for the maintenance of pregnancy. It is secreted from the corpus luteum in response to luteinizing hormone (LH). The primary contraceptive effect of exogenous progestins involves the suppression of the midcycle surge of LH. At the cellular level, progestins diffuse freely into target cells and bind to the progesterone receptor. Target cells include the female reproductive tract, the mammary gland, the hypothalamus, and the pituitary. Once bound to the receptor, progestins slow the frequency of release of gonadotropin releasing hormone from the hypothalamus and blunt the pre-ovulatory LH surge, thereby preventing follicular maturation and ovulation. Progesterone has minimal estrogenic and androgenic activity. Progesterone is metabolized hepatically to pregnanediol and conjugated with glucuronic acid.

Medoxyprogesterone (MAH), also known as 6 α -methyl-17-hydroxyprogesterone, is a synthetic progestin with a pharmacological activity about 15 times greater than progesterone. MAH is used for the treatment of renal and endometrial carcinomas, amenorrhea, abnormal uterine bleeding, and endometriosis associated with hormonal imbalance. MAH has a stimulatory effect on respiratory centers and has been used in cases of low blood oxygenation caused by sleep apnea, chronic obstructive pulmonary disease, or hypercapnia.

Mifepristone, also known as RU-486, is an antiprogestrone drug that blocks receptors of progesterone. It counteracts the effects of progesterone, which is needed to sustain pregnancy. Mifepristone induces spontaneous abortion when administered in early pregnancy followed by treatment with the prostaglandin misoprostol. Further studies show that mifepristone at a

substantially lower dose can be highly effective as a postcoital contraceptive when administered within five days after unprotected intercourse, thus providing women with a "morning-after pill" in case of contraceptive failure or sexual assault. Mifepristone also has potential uses in the treatment of breast and ovarian cancers in cases in which tumors are progesterone-dependent. It interferes with steroid-dependent growth of brain meningiomas, and may be useful in treatment of endometriosis where it blocks the estrogen-dependent growth of endometrial tissues. It may also be useful in treatment of uterine fibroid tumors and Cushing's Syndrome. Mifepristone binds to glucocorticoid receptors and interferes with cortisol binding. Mifepristone also may act as an anti-glucocorticoid and be effective for treating conditions where cortisol levels are elevated such as AIDS, anorexia nervosa, ulcers, diabetes, Parkinson's disease, multiple sclerosis, and Alzheimer's disease.

Danazol is a synthetic steroid derived from ethinyl testosterone. Danazol indirectly reduces estrogen production by lowering pituitary synthesis of follicle-stimulating hormone and LH. Danazol also binds to sex hormone receptors in target tissues, thereby exhibiting anabolic, antiestrogenic, and weakly androgenic activity. Danazol does not possess any progestogenic activity, and does not suppress normal pituitary release of corticotropin or release of cortisol by the adrenal glands. Danazol is used in the treatment of endometriosis to relieve pain and inhibit endometrial cell growth. It is also used to treat fibrocystic breast disease and hereditary angioedema.

Corticosteroids are used to relieve inflammation and to suppress the immune response. They inhibit eosinophil, basophil, and airway epithelial cell function by regulation of cytokines that mediate the inflammatory response. They inhibit leukocyte infiltration at the site of inflammation, interfere in the function of mediators of the inflammatory response, and suppress the humoral immune response. Corticosteroids are used to treat allergies, asthma, arthritis, and skin conditions. Beclomethasone is a synthetic glucocorticoid that is used to treat steroid-dependent asthma, to relieve symptoms associated with allergic or nonallergic (vasomotor) rhinitis, or to prevent recurrent nasal polyps following surgical removal. The anti-inflammatory and vasoconstrictive effects of intranasal beclomethasone are 5000 times greater than those produced by hydrocortisone.

Budesonide is a corticosteroid used to control symptoms associated with allergic rhinitis or asthma. Budesonide has high topical anti-inflammatory activity but low systemic activity. Dexamethasone is a synthetic glucocorticoid used in anti-inflammatory or immunosuppressive compositions. It is also used in inhalants to prevent symptoms of asthma. Due to its greater ability to reach the central nervous system, dexamethasone is usually the treatment of choice to control cerebral edema.

Dexamethasone is approximately 20-30 times more potent than hydrocortisone and 5-7 times more potent than prednisone. Prednisone is metabolized in the liver to its active form, prednisolone, a glucocorticoid with anti-inflammatory properties. Prednisone is approximately 4 times more potent than hydrocortisone and the duration of action of prednisone is intermediate between hydrocortisone

and dexamethasone. Prednisone is used to treat allograft rejection, asthma, systemic lupus erythematosus, arthritis, ulcerative colitis, and other inflammatory conditions. Betamethasone is a synthetic glucocorticoid with anti-inflammatory and immunosuppressive activity and is used to treat psoriasis and fungal infections, such as athlete's foot and ringworm.

5 The anti-inflammatory actions of corticosteroids are thought to involve phospholipase A₂ inhibitory proteins, collectively called lipocortins. Lipocortins, in turn, control the biosynthesis of potent mediators of inflammation such as prostaglandins and leukotrienes by inhibiting the release of the precursor molecule arachidonic acid. Proposed mechanisms of action include decreased IgE synthesis, increased number of β-adrenergic receptors on leukocytes, and decreased arachidonic acid 10 metabolism. During an immediate allergic reaction, such as in chronic bronchial asthma, allergens bridge the IgE antibodies on the surface of mast cells, which triggers these cells to release chemotactic substances. Mast cell influx and activation, therefore, is partially responsible for the inflammation and hyperirritability of the oral mucosa in asthmatic patients. This inflammation can 15 be retarded by administration of corticosteroids.

15 The metabolism of many drug substances, such as acetaminophen, in the liver leads to the generation of catechol (ortho-hydroxy benzene) containing intermediates. Catechol-containing intermediates may themselves be hepatotoxic by generating reactive intermediates or they may be further metabolized for example via catechol-o-methyl transferase to generate reactive methylating agents. As such the contribution of catechol-containing compounds to cell toxicity is of interest.

20 The effects upon liver metabolism and hormone clearance mechanisms are important to understand the pharmacodynamics of a drug. For example, the human C3A cell line is a clonal derivative of HepG2/C3 (hepatoma cell line, isolated from a 15-year-old male with liver tumor), which was selected for strong contact inhibition of growth. The use of a clonal population enhances the reproducibility of the cells. C3A cells have many characteristics of primary human hepatocytes 25 in culture: i) expression of insulin receptor and insulin-like growth factor II receptor; ii) secretion of a high ratio of serum albumin compared with α-fetoprotein; iii) conversion of ammonia to urea and glutamine; iv) metabolism of aromatic amino acids; and v) proliferation in glucose-free and insulin-free medium. The C3A cell line is now well established as an *in vitro* model of the mature human liver (Mickelson, J.K. et al. (1995) Hepatology 22:866-875; Nagendra, A.R. et al. (1997) Am. J. 30 Physiol. 272:G408-G416).

Anti-lipemic Agents

Gemfibrozil is an anti-lipemic agent that lowers serum triglycerides and produces favorable changes in lipoproteins. Gemfibrozil is effective in reducing the risk of coronary heart disease in men (Frick, M.H., et al. (1987) New Engl. J. Med. 317:1237-1245). The compound can inhibit 35 peripheral lipolysis and decrease hepatic extraction of free fatty acids, which decreases hepatic

triglyceride production. Gemfibrozil also inhibits the synthesis and increases the clearance of apolipoprotein B, a carrier molecule for VLDL. Gemfibrozil has variable effects on LDL cholesterol. Although it causes moderate reductions in patients with type IIa hyperlipoproteinemia, changes in patients with either type IIb or type IV hyperlipoproteinemia are unpredictable. In 5 general, the HMG-CoA reductase inhibitors are more effective than gemfibrozil in reducing LDL cholesterol. At the molecular level gemfibrozil may function as a peroxisome proliferator-activated receptor (PPAR) agonist. Gemfibrozil is rapidly and completely absorbed from the GI tract and undergoes enterohepatic recirculation. Gemfibrozil is metabolized by the liver and excreted by the kidneys, mainly as metabolites, one of which possesses pharmacologic activity. Gemfibrozil causes 10 peroxisome proliferation and hepatocarcinogenesis in rats, which is a cause for concern generally for fibric acid derivative drugs. In humans, fibric acid derivatives are known to increase the risk of gall bladder disease although gemfibrozil is better tolerated than other fibrates. The relative safety of gemfibrozil in humans compared to rodent species including rats may be attributed to differences in metabolism and clearance of the compound in different species (Dix, K.J. et al. (1999) Drug Metab. 15 Distrib. 27:138-146; Thomas, B.F. et al. (1999) Drug Metab. Distrib. 27:147-157).

Fibric acid compounds are usually well tolerated. Side effects may occur in 5-10% of patients but most often are not sufficient to cause discontinuation of the drug. Gastrointestinal side effects are most common (up to 5% of patients). Other side effects are reported infrequently and include rash, urticaria, hair loss, myalgias, fatigue, headache, impotence, and anemia. Minor 20 increases in liver transaminases and decreases in alkaline phosphatase have been reported. Clofibrate and bezafibrate have been reported to potentiate the action of oral anticoagulants, in part by displacing them from their binding sites on albumin. A myositis-flu-like syndrome occasionally occurs in subjects on gemfibrozil and may occur in up to 5% of patients treated with a combination of an HMG CoA reductase inhibitor and gemfibrozil, particularly when higher doses of the reductase 25 inhibitor are used. Clofibrate, and indeed all the fibrates, increase the lithogenicity of bile. Renal failure is a relative contraindication to the use of fibric acid agents, as is hepatic dysfunction. Gemfibrozil is used to treat the hyperlipidemia of renal failure, but it should be used for this application with caution and at a reduced dosage. These agents should not be used during pregnancy or in children. Clofibrate is available for oral administration. The usual dose is 2 g/day in divided 30 doses. This compound is rarely used today, but it may be useful in patients with dysbetalipoproteinemia who do not respond to gemfibrozil.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative 35 disorders, and cancers.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, kinases and phosphatases, referred to collectively as 'KPP' and individually as 'KPP-1,' 'KPP-2,' 'KPP-3,' 'KPP-4,' 'KPP-5,' 'KPP-6,' 'KPP-7,' 'KPP-8,' 'KPP-9,' 'KPP-10,' 'KPP-11,' 'KPP-12,' 'KPP-13,' 5 'KPP-14,' 'KPP-15,' 'KPP-16,' 'KPP-17,' 'KPP-18,' 'KPP-19,' 'KPP-20,' 'KPP-21,' 'KPP-22,' 'KPP-23,' 'KPP-24,' 'KPP-25,' 'KPP-26,' 'KPP-27,' 'KPP-28,' 'KPP-29,' 'KPP-30,' 'KPP-31,' 'KPP-32,' 'KPP-33,' 'KPP-34,' 'KPP-35,' 'KPP-36,' 'KPP-37,' 'KPP-38,' 'KPP-39,' 'KPP-40,' 'KPP-41,' 'KPP-42,' 'KPP-43,' 'KPP-44,' 'KPP-45,' 'KPP-46,' 'KPP-47,' 'KPP-48,' 'KPP-49,' 'KPP-50,' 'KPP-51,' 'KPP-52,' 'KPP-53,' 'KPP-54,' 'KPP-55,' and 'KPP-56' and methods for using 10 these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related 15 embodiments provide methods for utilizing the purified kinases and phosphatases and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID 20 NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-56.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a 25 polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-56. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:57-112.

35 Still another embodiment provides a recombinant polynucleotide comprising a promoter

sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at

least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, b) 5 a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides 10 comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, 15 the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, b) 20 a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target 25 polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a 30 naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and a pharmaceutically acceptable excipient. 35 In one embodiment, the composition can comprise an amino acid sequence selected from the group

consisting of SEQ ID NO:1-56. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as
5 an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group
10 consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. The method comprises a) contacting a sample comprising the polypeptide with a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment
15 provides a method of treating a disease or condition associated with decreased expression of functional KPP, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide
20 comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. The method comprises a)
25 contacting a sample comprising the polypeptide with a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional KPP, comprising administering to a patient in need of such treatment the composition.

30 Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active
35 fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ

ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. The method comprises a) combining the polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the
5 polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an
10 amino acid sequence selected from the group consisting of SEQ ID NO:1-56, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the
15 polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, the method comprising a) contacting a sample comprising the target polynucleotide with a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the
25 target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least
30 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide
35 complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization

occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

15 BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

25 Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

30 Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

35 DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of
5 the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

10 Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of
15 describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

20 "KPP" refers to the amino acid sequences of substantially purified KPP obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of KPP. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other
25 compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

An "allelic variant" is an alternative form of the gene encoding KPP. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or
30 many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding KPP include those sequences with deletions,
35 insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as KPP or a

polypeptide with at least one functional characteristic of KPP. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding KPP, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide encoding KPP. The 5 encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent KPP. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of KPP is retained. For example, negatively 10 charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

15 The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein 20 molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological 25 activity of KPP. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of KPP either by directly interacting with KPP or by acting on components of the biological pathway in which KPP participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments 30 thereof, such as Fab, F(ab')₂, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind KPP polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. 35 Commonly used carriers that are chemically coupled to peptides include bovine serum albumin,

thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that makes contact with a particular antibody. When a protein or a fragment of a protein is used to 5 immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a 10 specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include 15 deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH₂), which may improve a desired property, e.g., 20 resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a 25 vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left- 30 handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" 35 (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'- deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis

or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

5 The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic KPP, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

10 "Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

15 A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding KPP or fragments of KPP may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's 20 solution, dry milk, salmon sperm DNA, etc.).

25 "Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

30 "Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His

	Asp	Asn, Glu
	Cys	Ala, Ser
	Gln	Asn, Glu, His
	Glu	Asp, Gln, His
5	Gly	Ala
	His	Asn, Arg, Gln, Glu
	Ile	Leu, Val
	Leu	Ile, Val
	Lys	Arg, Gln, Glu
10	Met	Leu, Ile
	Phe	His, Met, Leu, Trp, Tyr
	Ser	Cys, Thr
	Thr	Ser, Val
15	Trp	Phe, Tyr
	Tyr	His, Phe, Trp
	Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation,
 20 (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide.

25 Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

30 A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a
 35 diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

40 A "fragment" is a unique portion of KPP or a polynucleotide encoding KPP which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise

up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may 5 be encompassed by the present embodiments.

10

A fragment of SEQ ID NO:57-112 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:57-112, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:57-112 can be employed in one or more embodiments of methods of the invention, for example, in hybridization 15 and amplification technologies and in analogous methods that distinguish SEQ ID NO:57-112 from related polynucleotides. The precise length of a fragment of SEQ ID NO:57-112 and the region of SEQ ID NO:57-112 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-56 is encoded by a fragment of SEQ ID NO:57-112. A 20 fragment of SEQ ID NO:1-56 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-56. For example, a fragment of SEQ ID NO:1-56 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-56. The precise length of a fragment of SEQ ID NO:1-56 and the region of SEQ ID NO:1-56 to which the fragment corresponds can be determined based on the intended purpose for the fragment 25 using one or more analytical methods described herein or otherwise known in the art.

A “full length” polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A “full length” polynucleotide sequence encodes a “full length” polypeptide sequence.

“Homology” refers to sequence similarity or, alternatively, sequence identity, between two 30 or more polynucleotide sequences or two or more polypeptide sequences.

The terms “percent identity” and “% identity,” as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two 35 sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part 5 of the LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as 10 the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on 15 the Internet at ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 20 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

25 *Reward for match: 1*

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

30 *Word Size: 11*

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at 35 least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous

nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

30 *Open Gap: 11 and Extension Gap: 1 penalties*

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

35 Percent identity may be measured over the length of an entire defined polypeptide sequence,

for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment 5 length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

“Human artificial chromosomes” (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

10 The term “humanized antibody” refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

15 “Hybridization” refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the “washing” step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly 20 matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 µg/ml sheared, denatured salmon 25 sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of 30 the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present 35 invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS,

for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 µg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C₀t or R₀t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of KPP which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of KPP which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of KPP. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of KPP.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide,

oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

5 "Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

10 "Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

15 "Post-translational modification" of an KPP may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of KPP.

20 "Probe" refers to nucleic acids encoding KPP, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a 25 DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 30 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, 35 Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in

Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research,

5 Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to 100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer

10 selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for

15 Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping

20 Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in

25 hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of 30 sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid 35 sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a

vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be used to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

5 A “regulatory element” refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription, translation, or RNA stability.

10 “Reporter molecules” are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

15 An “RNA equivalent,” in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

20 The term “sample” is used in its broadest sense. A sample suspected of containing KPP, nucleic acids encoding KPP, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

25 The terms “specific binding” and “specifically binding” refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope “A,” the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

30 The term “substantially purified” refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A “substitution” refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

35 “Substrate” refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers,

microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

5 "Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to,
10 bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

15 A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by
20 infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria,
25 cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

30 A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least
35 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least

99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human kinases and phosphatases (KPP), the polynucleotides encoding KPP, and the use of these compositions for the diagnosis, treatment, or prevention of cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode

polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers (PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows amino acid residues comprising signature sequences, domains, motifs, potential phosphorylation sites, and potential glycosylation sites. Column 5 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are kinases and phosphatases. For example, SEQ ID NO:3 is 97% identical, from residue L85 to residue C198, to rat protein tyrosine phosphatase (GenBank ID g409023) as determined by the Basic Local Alignment Search Tool (BLAST). (See Table 2.) The BLAST probability score is 2.4e-94, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:3 has homology to dual specificity phosphatase 1, which is a stress factor-induced, non-membrane spanning protein phosphatase that inactivates members of the mitogen activated protein kinase family, is involved in chemotaxis, cell proliferation, and stress responses, and may possibly be involved in apoptosis and heat shock, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:3 also contains a dual specificity phosphatase, catalytic domain and a rhodanese-like homology domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM/SMART databases of conserved protein families/domains. (See Table 3.) Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:3 is a protein tyrosine phosphatase.

In an alternative example, SEQ ID NO:26 is 98% identical, from residue H207 to residue W902, to human leucine zipper bearing kinase (GenBank ID g2879898) as determined by BLAST. The BLAST probability score is 0.0. SEQ ID NO:26 also has homology to proteins that are localized to the cytoplasm, nucleus or golgi, have kinase activity, and are protein kinases, as 5 determined by BLAST analysis using the PROTEOME database. SEQ ID NO:26 also contains a serine/threonine protein kinase catalytic domain, a tyrosine kinase catalytic domain, and a protein kinase domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based SMART and PFAM databases of conserved protein families/domains. Data from BLIMPS analyses and BLAST analyses against the PRODOM and DOMO databases provide 10 further corroborative evidence that SEQ ID NO:26 is a leucine zipper bearing protein kinase.

In an alternative example, SEQ ID NO:35 is 99% identical, from residue M1 to residue V552, to human protein kinase C-theta (GenBank ID g558099) as determined by BLAST. The BLAST probability score is 0.0. SEQ ID NO:35 also has homology to protein kinase C-theta which is involved in T cell activation and protection from apoptosis, may play a role in insulin and 15 multidrug resistance, and may regulate mitosis, preadipocyte differentiation, and spermatogenesis; further, rat Pkcq may play a role in hyperglycemia and hypertriglyceridemia, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:35 also contains protein kinase C conserved region 1 (C1), extension to Ser/Thr-type protein kinases, serine/threonine protein kinases catalytic, phorbol esters/diacylglycerol binding, protein kinase, and protein kinase C terminal 20 domains as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM/SMART databases of conserved protein families/domains. Data from BLIMPS, MOTIFS, and PROFILESCAN analyses, and BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:35 is a protein kinase C-theta.

25 In an alternative example, SEQ ID NO:49 is 98% identical, from residue M1 to residue W1230, to mouse NIK (GenBank ID g1872546) as determined by the BLAST. The BLAST probability score is 0.0, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:49 also has homology to proteins that are related to the Ste20 family of kinases, that bind Nck and MEKK1, that activate the JNK signaling pathway, that 30 may be involved in TNF- α signaling, and that, if mutated, prevent development of somites or hindgut, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:49 also contains a domain found in NIK1-like kinases, mouse citron and yeast ROM1 and ROM2, and also a serine/threonine kinase catalytic domain, as determined by searching for statistically significant 35 matches in the hidden Markov model (HMM)-based PFAM and SMART database of conserved protein families/domains. (See Table 3.) Data from MOTIFS and PROFILESCAN analyses, and

BLAST analyses against the PRODOM and DOMO databases, provide further corroborative evidence that SEQ ID NO:49 is a serine/threonine kinase. SEQ ID NO:1-2, SEQ ID NO:4-25, SEQ ID NO:27-34, SEQ ID NO:36-48, and SEQ ID NO:50-56 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-56 are described in Table

5 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5') and stop (3') positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:57-112 or that distinguish between SEQ ID NO:57-112 and related 10 15 polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, 20 25 the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_XXXXXX_N₁_N₂_YYYYY_N₃_N₄ represents a "stitched" sequence in which XXXXXX is the identification number of the cluster of sequences to which the algorithm was applied, and YYYYY is 30 35 the number of the prediction generated by the algorithm, and N_{1,2,3...}, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FLXXXXXX_gAAAAAA_gBBBBB_1_N is a "stretched" sequence, with XXXXXX being the Incyte project identification number, gAAAAAA being the GenBank identification number of the human

genomic sequence to which the "exon-stretching" algorithm was applied, *gBBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier 5 (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, *gBBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

	Prefix	Type of analysis and/or examples of programs
10	GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
	GBI	Hand-edited analysis of genomic sequences.
	FL	Stitched or stretched genomic sequences (see Example V).
15	INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

20 Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

25 Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the 30 identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence.

Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while 5 n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses KPP variants. Various embodiments of KPP variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the KPP amino acid sequence, and can contain at least one functional or structural characteristic of KPP.

10 Various embodiments also encompass polynucleotides which encode KPP. In a particular embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:57-112, which encodes KPP. The polynucleotide sequences of SEQ ID NO:57-112, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the 15 sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding KPP. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding KPP. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence 20 selected from the group consisting of SEQ ID NO:57-112 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:57-112. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

25 In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding KPP. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding KPP, but will generally have a greater or lesser number of nucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a 30 polynucleotide encoding KPP over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding KPP. For example, a polynucleotide comprising a sequence of SEQ ID NO:76, a polynucleotide 35 comprising a sequence of SEQ ID NO:92, and a polynucleotide comprising a sequence of SEQ ID

NO:96 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:79 and a polynucleotide comprising a sequence of SEQ ID NO:80 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:84 and a polynucleotide comprising a sequence of SEQ ID NO:94 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:88 and a polynucleotide comprising a sequence of SEQ ID NO:99 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:89 and a polynucleotide comprising a sequence of SEQ ID NO:97 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:93 and a polynucleotide comprising a sequence of SEQ ID NO:100 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:98 and a polynucleotide comprising a sequence of SEQ ID NO:111 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:104 and a polynucleotide comprising a sequence of SEQ ID NO:105 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of KPP.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding KPP, some bearing minimal similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring KPP, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode KPP and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring KPP under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding KPP or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding KPP and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode KPP and KPP derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems

using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding KPP or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown 5 in SEQ ID NO:57-112 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A.R. (1987) *Methods Enzymol.* 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of 10 the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines 15 such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems). Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which 20 are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding KPP may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as 25 promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) *PCR Methods Applic.* 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known 30 genomic locus and surrounding sequences (Triglia, T. et al. (1988) *Nucleic Acids Res.* 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known 35 sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) *PCR Methods Applic.* 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) *Nucleic Acids Res.* 19:3055-3060). Additionally, one

may use PCR, nested primers, and PROMOTERFINDER libraries (BD Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth 5 MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) 10 library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments 15 which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode KPP may be cloned in recombinant DNA molecules that direct expression of KPP, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent 20 polypeptides may be produced and used to express KPP.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter KPP-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be 30 used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 35 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat.

Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of KPP, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then 5 subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are 10 optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding KPP may be synthesized, in whole or in 15 part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232). Alternatively, KPP itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase 20 techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of KPP, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

25 The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active KPP, the polynucleotides encoding KPP or 30 derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in 35 polynucleotides encoding KPP. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding

KPP. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding KPP and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) *Results Probl. Cell Differ.* 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding KPP and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding KPP. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509; Engelhard, E.K. et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3224-3227; Sandig, V. et al. (1996) *Hum. Gene Ther.* 7:1937-1945; Takamatsu, N. (1987) *EMBO J.* 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659; Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) *Cancer Gen. Ther.* 5:350-356; Yu, M. et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6340-6344; Buller, R.M. et al. (1985) *Nature* 317:813-815; McGregor, D.P. et al. (1994) *Mol. Immunol.* 31:219-226; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding KPP. For example, routine cloning, subcloning, and propagation of polynucleotides encoding KPP can be achieved using a multifunctional *E. coli*

vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSSPORT1 plasmid (Invitrogen).

Ligation of polynucleotides encoding KPP into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro*

5 transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) *J. Biol. Chem.* 264:5503-5509). When large quantities of KPP are needed, e.g. for the production of antibodies, vectors which direct high level expression of KPP may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

10 Yeast expression systems may be used for production of KPP. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation

15 (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) *Methods Enzymol.* 153:516-544; Scorer, C.A. et al. (1994) *Bio/Technology* 12:181-184).

Plant systems may also be used for expression of KPP. Transcription of polynucleotides encoding KPP may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) *EMBO J.* 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) *EMBO J.* 3:1671-1680; Broglie, R. et al. (1984) *Science* 224:838-843; Winter, J. et al. (1991) *Results Probl. Cell Differ.* 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (*The McGraw Hill Yearbook of Science and Technology* (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding KPP may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses KPP in host cells (Logan, J. and T. Shenk (1984) *Proc. Natl. Acad. Sci. USA* 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are

constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) *Nat. Genet.* 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression 5 of KPP in cell lines is preferred. For example, polynucleotides encoding KPP can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to 10 confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine 15 phosphoribosyltransferase genes, for use in *tk* and *apr* cells, respectively (Wigler, M. et al. (1977) *Cell* 11:223-232; Lowy, I. et al. (1980) *Cell* 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et 20 al. (1980) *Proc. Natl. Acad. Sci. USA* 77:3567-3570; Colbere-Garapin, F. et al. (1981) *J. Mol. Biol.* 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) *Proc. Natl. Acad. Sci. USA* 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; BD Clontech), β -glucuronidase and its substrate β -glucuronide, or luciferase and its substrate luciferin may be used. 25 These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) *Methods Mol. Biol.* 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if 30 the sequence encoding KPP is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding KPP can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding KPP under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

35 In general, host cells that contain the polynucleotide encoding KPP and that express KPP

may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

5 Immunological methods for detecting and measuring the expression of KPP using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on KPP is preferred, but a
10 competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

15 A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding KPP include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding KPP, or any fragments thereof, may be cloned
20 into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels
25 which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding KPP may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a
30 transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode KPP may be designed to contain signal sequences which direct secretion of KPP through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the
35 inserted polynucleotides or to process the expressed protein in the desired fashion. Such

modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding KPP may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric KPP protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of KPP activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the KPP encoding sequence and the heterologous protein sequence, so that KPP may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled KPP may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ³⁵S-methionine.

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that specifically bind to KPP. One or more test compounds may be screened for specific binding to KPP. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to KPP. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of KPP can be used to screen for binding of test compounds, such as antibodies, to KPP, a variant of KPP, or a combination of KPP and/or one or more variants KPP. In an embodiment, a variant of KPP can be used to screen for compounds that bind to a variant of KPP, but not to KPP having the exact sequence of a sequence of SEQ ID

5 NO:1-56. KPP variants used to perform such screening can have a range of about 50% to about 99% sequence identity to KPP, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to KPP can be closely related to the natural ligand of KPP, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor KPP (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to KPP can be closely related to the natural receptor to which KPP binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for KPP which is capable of propagating a signal, or a decoy receptor for KPP which is not capable of propagating a signal (Ashkenazi, A. and V.M. Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG₁ (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

25 In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of KPP. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of KPP. In another 30 embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of KPP.

35 In an embodiment, anticalins can be screened for specific binding to KPP, fragments of KPP, or variants of KPP. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A.

(2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made 5 using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit KPP involves producing appropriate cells which express KPP, either as a secreted protein or 10 on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing KPP or cell membrane fractions which contain KPP are then contacted with a test compound and binding, stimulation, or inhibition of activity of either KPP or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is 15 detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with KPP, either in solution or affixed to a solid support, and detecting the binding of KPP to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical 20 libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 25 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells 30 (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

KPP, fragments of KPP, or variants of KPP may be used to screen for compounds that modulate the activity of KPP. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for KPP 35 activity, wherein KPP is combined with at least one test compound, and the activity of KPP in the

presence of a test compound is compared with the activity of KPP in the absence of the test compound. A change in the activity of KPP in the presence of the test compound is indicative of a compound that modulates the activity of KPP. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising KPP under conditions suitable for KPP activity, and the assay 5 is performed. In either of these assays, a test compound which modulates the activity of KPP may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding KPP or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem 10 (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, 15 M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell 20 blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding KPP may also be manipulated *in vitro* in ES cells derived from 25 human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding KPP can also be used to create "knockin" humanized animals 30 (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding KPP is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. 35 Alternatively, a mammal inbred to overexpress KPP, e.g., by secreting KPP in its milk, may also

serve as a convenient source of that protein (Janne, J. et al. (1998) *Biotechnol. Annu. Rev.* 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of KPP and kinases and phosphatases. In addition, examples of tissues expressing KPP can be found in Table 6 and can also be found in Example XI. Therefore, KPP appears to play a role in cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers. In the treatment of disorders associated with increased KPP expression or activity, it is desirable to decrease the expression or activity of KPP. In the treatment of disorders associated with decreased KPP expression or activity, it is desirable to increase the expression or activity of KPP.

Therefore, in one embodiment, KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome,

allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with 5 lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, 10 systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, 15 amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal 20 familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral 25 nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial 30 frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR 35 syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-

Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis, hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis, carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipoid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia, hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias such as multiple myeloma, and lymphomas such as Hodgkin's disease.

In another embodiment, a vector capable of expressing KPP or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified KPP in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of KPP may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of KPP including, but not limited to, those listed above.

In a further embodiment, an antagonist of KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP. Examples of such

disorders include, but are not limited to, those cardiovascular diseases, immune system disorders, neurological disorders, disorders affecting growth and development, lipid disorders, cell proliferative disorders, and cancers described above. In one aspect, an antibody which specifically binds KPP may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing 5 a pharmaceutical agent to cells or tissues which express KPP.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding KPP may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of KPP including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, 10 or vector embodiments may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy 15 with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of KPP may be produced using methods which are generally known in the art. In particular, purified KPP may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind KPP. Antibodies to KPP may also be generated using methods that are well known in the art. Such antibodies may include, but are not 20 limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muylldermans, S. (2001) J. 25 Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with KPP or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are 30 not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lyssolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to KPP 35 have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of

at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of KPP amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

5 Monoclonal antibodies to KPP may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) *Nature* 256:495-497; Kozbor, D. et al. (1985) *J. Immunol. Methods* 81:31-42; Cote, R.J. et al. (1983) *Proc. Natl. Acad. Sci. USA* 80:2026-2030;

10 Cole, S.P. et al. (1984) *Mol. Cell Biol.* 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) *Proc. Natl. Acad. Sci. USA* 81:6851-6855; Neuberger, M.S. et al. (1984) *Nature* 312:604-608; Takeda, S. et al. (1985) *Nature* 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce KPP-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) *Proc. Natl. Acad. Sci. USA* 88:10134-10137).

20 Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:3833-3837; Winter, G. et al. (1991) *Nature* 349:293-299).

Antibody fragments which contain specific binding sites for KPP may also be generated. 25 For example, such fragments include, but are not limited to, F(ab')₂ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')₂ fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) *Science* 246:1275-1281).

30 Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between KPP and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies 35 reactive to two non-interfering KPP epitopes is generally used, but a competitive binding assay may

also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for KPP. Affinity is expressed as an association constant, K_a , which is defined as the molar concentration of KPP-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions.

5 The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple KPP epitopes, represents the average affinity, or avidity, of the antibodies for KPP. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular KPP epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the KPP-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of KPP, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; 10 Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

15

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, 20 preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of KPP-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding KPP, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding KPP. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding 25 KPP (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. 30 et al. (1998) *J. Allergy Clin. Immunol.* 102:469-475; Scanlon, K.J. et al. (1995) *FASEB J.* 9:1288-

1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) *Blood* 76:271-278; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) *Pharmacol. Ther.* 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems 5 known in the art (Rossi, J.J. (1995) *Br. Med. Bull.* 51:217-225; Boado, R.J. et al. (1998) *J. Pharm. Sci.* 87:1308-1315; Morris, M.C. et al. (1997) *Nucleic Acids Res.* 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding KPP may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) *Science* 288:669-672), severe combined immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) *Science* 270:475-480; Bordignon, C. et al. (1995) *Science* 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) *Cell* 75:207-216; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:643-666; Crystal, R.G. et al. (1995) *Hum. Gene Therapy* 6:667-703), thalassamias, 10 familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) *Science* 270:404-410; Verma, I.M. and N. Somia (1997) *Nature* 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular 15 parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) *Nature* 335:395-396; Poeschla, E. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in KPP expression or regulation causes 20 disease, the expression of KPP from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in KPP are treated by constructing mammalian expression vectors encoding KPP and introducing these vectors by mechanical means into KPP-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic 25 gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) *Annu. Rev. Biochem.* 62:191-217; Ivics, Z. (1997) *Cell* 91:501-510; Boulay, J.-L. and H. Récipon (1998) *Curr. Opin. Biotechnol.* 9:445-450).

Expression vectors that may be effective for the expression of KPP include, but are not 35 limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors

(Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (BD Clontech, Palo Alto CA). KPP may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β -actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*), or (iii) a tissue-specific promoter or the native promoter of the endogenous gene encoding KPP from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to KPP expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding KPP under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., CD4 $^{+}$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in

the art of gene therapy and have been well documented (Ranga, U. et al. (1997) *J. Virol.* 71:7020-7029; Bauer, G. et al. (1997) *Blood* 89:2259-2267; Bonyhadi, M.L. (1997) *J. Virol.* 71:4707-4716; Ranga, U. et al. (1998) *Proc. Natl. Acad. Sci. USA* 95:1201-1206; Su, L. (1997) *Blood* 89:2283-2290).

5 In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding KPP to cells which have one or more genetic abnormalities with respect to the expression of KPP. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas
10 (Csete, M.E. et al. (1995) *Transplantation* 27:263-268). Potentially useful adenoviral vectors are described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; *Annu. Rev. Nutr.* 19:511-544) and Verma, I.M. and N. Somia (1997; *Nature* 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver 15 polynucleotides encoding KPP to target cells which have one or more genetic abnormalities with respect to the expression of KPP. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing KPP to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector
20 has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) *Exp. Eye Res.* 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under
25 the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; *J. Virol.* 73:519-532) and Xu, H. et al. (1994; *Dev. Biol.* 163:152-161). The manipulation of cloned herpesvirus sequences, the generation
30 of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding KPP to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based
35 on the SFV genome (Garoff, H. and K.-J. Li (1998) *Curr. Opin. Biotechnol.* 9:464-469). During

alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for KPP into the 5 alphavirus genome in place of the capsid-coding region results in the production of a large number of KPP-coding RNAs and the synthesis of high levels of KPP in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy 10 application (Dryga, S.A. et al. (1997) *Virology* 228:74-83). The wide host range of alphaviruses will allow the introduction of KPP into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA 15 transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of 20 polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

25 Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding KPP.

30 Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of 35 candidate targets may also be evaluated by testing accessibility to hybridization with complementary

oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis.

5 Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding KPP. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

10 RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as
15 inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytosine, guanine, thymine, and uracil which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or
25 DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as
30 siRNA. siRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. siRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

35 siRNA can be generated indirectly by introduction of dsRNA into the targeted cell.

Alternatively, siRNA can be synthesized directly and introduced into a cell by transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable siRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target sequences with significant homology to other coding sequences can be eliminated from consideration. The selected siRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene can be determined, for example, by northern analysis methods using the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined, for example, by microarray methods; by polyacrylamide gel electrophoresis; and by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding KPP. Compounds

which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective 5 compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased KPP expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding KPP may be therapeutically useful, and in the treatment of disorders associated with decreased KPP expression or activity, a compound which specifically promotes expression of the 10 polynucleotide encoding KPP may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in altering polynucleotide expression; selection from an existing, commercially-available or 15 proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding KPP is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted 20 biochemical system. Alterations in the expression of a polynucleotide encoding KPP are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding KPP. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to 25 one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) 30 Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, 35 T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) *Nat. Biotechnol.* 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of KPP, antibodies to KPP, and mimetics, agonists, antagonists, or inhibitors of KPP.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising KPP or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, KPP or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to

transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) *Science* 285:1569-1572).

. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, 5 monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example KPP or fragments thereof, antibodies of KPP, and agonists, antagonists or inhibitors of KPP, which 10 ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED₅₀ (the dose therapeutically effective in 50% of the population) or LD₅₀ (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as the LD₅₀/ED₅₀ ratio. Compositions which exhibit large 15 therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED₅₀ with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

20 The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and 25 response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 µg to 100,000 µg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. 30 Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind KPP may be used for the 35 diagnosis of disorders characterized by expression of KPP, or in assays to monitor patients being

treated with KPP or agonists, antagonists, or inhibitors of KPP. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for KPP include methods which utilize the antibody and a label to detect KPP in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, 5 and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring KPP, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of KPP expression. Normal or standard values for KPP expression are established by combining body fluids or cell extracts 10 taken from normal mammalian subjects, for example, human subjects, with antibodies to KPP under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of KPP expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

15 In another embodiment of the invention, polynucleotides encoding KPP may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of KPP may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of 20 KPP, and to monitor regulation of KPP levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting 25 polynucleotides, including genomic sequences, encoding KPP or closely related molecules may be used to identify nucleic acid sequences which encode KPP. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding KPP, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the KPP encoding sequences. The hybridization probes of the subject 30 invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:57-112 or from genomic sequences including promoters, enhancers, and introns of the KPP gene.

Means for producing specific hybridization probes for polynucleotides encoding KPP 35 include the cloning of polynucleotides encoding KPP or KPP derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA

polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S , or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding KPP may be used for the diagnosis of disorders associated with

5 expression of KPP. Examples of such disorders include, but are not limited to, a cardiovascular disease such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris,

10 myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of

15 cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse

20 interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural

25 effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an immune system disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal

30 dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation,

35 osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid

arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, viral, bacterial, fungal, parasitic, protozoal, and helminthic infections, and trauma; a neurological disorder such as epilepsy,

5 ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial

10 thrombophlebitis, myelitis and radiculitis, viral central nervous system disease, prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome, fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system including Down

15 syndrome, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis, inherited, metabolic, endocrine, and toxic myopathies, myasthenia gravis, periodic paralysis, mental disorders including mood, anxiety, and schizophrenic disorders, seasonal affective disorder (SAD), akathesia, amnesia, catatonia,

20 diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, Tourette's disorder, progressive supranuclear palsy, corticobasal degeneration, and familial frontotemporal dementia; a disorder affecting growth and development such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis,

25 primary thrombocythemia, renal tubular acidosis, anemia, Cushing's syndrome, achondroplastic dwarfism, Duchenne and Becker muscular dystrophy, epilepsy, gonadal dysgenesis, WAGR syndrome (Wilms' tumor, aniridia, genitourinary abnormalities, and mental retardation), Smith-Magenis syndrome, myelodysplastic syndrome, hereditary mucoepithelial dysplasia, hereditary keratodermas, hereditary neuropathies such as Charcot-Marie-Tooth disease and neurofibromatosis,

30 hypothyroidism, hydrocephalus, seizure disorders such as Sydenham's chorea and cerebral palsy, spina bifida, anencephaly, craniorachischisis, congenital glaucoma, cataract, and sensorineural hearing loss; a lipid disorder such as fatty liver, cholestasis, primary biliary cirrhosis; carnitine deficiency, carnitine palmitoyltransferase deficiency, myoadenylate deaminase deficiency, hypertriglyceridemia, lipid storage disorders such Fabry's disease, Gaucher's disease, Niemann-

35 Pick's disease, metachromatic leukodystrophy, adrenoleukodystrophy, GM₂ gangliosidosis, and

ceroid lipofuscinosis, abetalipoproteinemia, Tangier disease, hyperlipoproteinemia, diabetes mellitus, lipodystrophy, lipomatoses, acute panniculitis, disseminated fat necrosis, adiposis dolorosa, lipid adrenal hyperplasia, minimal change disease, lipomas, atherosclerosis, hypercholesterolemia, hypercholesterolemia with hypertriglyceridemia, primary hypoalphalipoproteinemia,

5 hypothyroidism, renal disease, liver disease, lecithin:cholesterol acyltransferase deficiency, cerebrotendinous xanthomatosis, sitosterolemia, hypocholesterolemia, Tay-Sachs disease, Sandhoff's disease, hyperlipidemia, hyperlipemia, lipid myopathies, and obesity; and a cell proliferative disorder such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal
10 hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia, and cancers including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, uterus, leukemias
15 such as multiple myeloma, and lymphomas such as Hodgkin's disease. Polynucleotides encoding KPP may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered KPP expression. Such qualitative or quantitative methods are well known in the art.

20 In a particular embodiment, polynucleotides encoding KPP may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding KPP may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and
25 compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding KPP in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

30 In order to provide a basis for the diagnosis of a disorder associated with expression of KPP, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding KPP, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects
35 with values from an experiment in which a known amount of a substantially purified polynucleotide

is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated,

5 hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or

10 overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

15 Additional diagnostic uses for oligonucleotides designed from the sequences encoding KPP may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding KPP, or a fragment of a polynucleotide complementary to the polynucleotide encoding KPP, and will be employed under optimized conditions for identification of
20 a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding KPP may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods
25 of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding KPP are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of
30 PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSSCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed *in silico* SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments
35 which assemble into a common consensus sequence. These computer-based methods filter out

sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

5 SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic
10 fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase
15 pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of KPP include radiolabeling or
20 biotinylation nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives
25 rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic
30 variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and
35 effective treatment regimen for that patient. For example, therapeutic agents which are highly

effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, KPP, fragments of KPP, or antibodies specific for KPP may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for

example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

5 In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared
10 with levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be
15 subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated
20 by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities
25 of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by
30 comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for KPP to quantify the levels of KPP expression. In one embodiment, the antibodies are used as elements on a microarray,
35 and protein expression levels are quantified by contacting the microarray with the sample and

detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) *Anal. Biochem.* 270:103-111; Mendoza, L.G. et al. (1999) *Biotechniques* 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at 5 each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) *Electrophoresis* 18:533-537), so proteome toxicant signatures may be 10 useful in the analysis of compounds which do not significantly affect the transcript image, but which alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological 15 sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing 20 the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of 25 protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, 30 T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) *Proc. Natl. Acad. Sci. USA* 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) *Proc. Natl. Acad. Sci. USA* 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays 35 are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding KPP may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members 5 of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) *Nat. Genet.* 10: 15:345-355; Price, C.M. (1993) *Blood Rev.* 7:127-134; Trask, B.J. (1991) *Trends Genet.* 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) *Proc. Natl. Acad. Sci. USA* 83:7353-7357).

15 Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding KPP on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of 20 DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information 25 is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) *Nature* 336:577-580). The nucleotide sequence of the 30 instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, KPP, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a 35 solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes

between KPP and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with KPP, or fragments thereof, and washed. Bound KPP is then detected by methods well known in the art. Purified KPP can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding KPP specifically compete with a test compound for binding KPP. In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with KPP.

In additional embodiments, the nucleotide sequences which encode KPP may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/423,226, U.S. Ser. No. 60/426,713, U.S. Ser. No. 60/429,766, and U.S. Ser. No. 60/447,043, are hereby expressly incorporated by reference.

25

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs are derived from cDNA libraries described in the LIFESEQ database (Incyte, Palo Alto CA). Some tissues are homogenized and lysed in guanidinium isothiocyanate, while others are homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates are centrifuged over CsCl cushions or extracted with chloroform. RNA is precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA are repeated as necessary to increase RNA purity. In some cases, RNA is treated with DNase. For most libraries, poly(A)+ RNA is isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN,

Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA is isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene is provided with RNA and constructs the corresponding cDNA

5 libraries. Otherwise, cDNA is synthesized and cDNA libraries are constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription is initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters are ligated to double stranded cDNA, and the cDNA is digested with the appropriate restriction enzyme or
10 enzymes. For most libraries, the cDNA is size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs are ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid
15 (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte, Palo Alto CA), pRARE (Incyte), or pINCY (Incyte), or derivatives thereof. Recombinant plasmids are transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5 α , DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

20 Plasmids obtained as described in Example I are recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids are purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the
25 R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids are resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA is amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps are carried out in a single reaction mixture. Samples are processed and stored in 384-well plates, and the concentration of amplified plasmid DNA is quantified fluorometrically using
30 PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSCAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II are sequenced as follows.
35 Sequencing reactions are processed using standard methods or high-throughput instrumentation such

as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions are prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the
5 ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides are carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading
10 frames within the cDNA sequences are identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences are selected for extension using the techniques disclosed in Example VIII.

Polynucleotide sequences derived from Incyte cDNAs are validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs
15 based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof are then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*,
20 *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary
25 structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries are performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences are assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) are used to extend Incyte cDNA
30 assemblages to full length. Assembly is performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages are screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences are translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide
35 sequences are subsequently analyzed by querying against databases such as the GenBank protein

databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes tools, programs, and algorithms used for the analysis and assembly of Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences are also used to identify polynucleotide sequence fragments from SEQ ID NO:57-112. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative kinases and phosphatases are initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once is set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode kinases and phosphatases, the encoded polypeptides are analyzed by querying against PFAM models for kinases and phosphatases. Potential kinases and phosphatases are also identified by homology to Incyte cDNA sequences that have been annotated as kinases and phosphatases. These selected Genscan-predicted sequences are then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences are then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence

predicted by Genscan, such as extra or omitted exons. BLAST analysis is also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage is available, this information is used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences are obtained by
5 assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences are derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

10 **"Stitched" Sequences**

Partial cDNA sequences are extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III are mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster is analyzed using an algorithm
15 based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that are subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval is present on more than one sequence in the cluster are identified, and intervals thus identified are considered to be equivalent by transitivity. For example, if an interval is present on a cDNA and two genomic
20 sequences, then all three intervals are considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified are then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants.
Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or
25 genomic sequence to genomic sequence) are given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences are translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan are corrected by comparison to the top BLAST hit from genpept. Sequences are further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

30 **"Stretched" Sequences**

Partial DNA sequences are extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III are queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog is then compared by BLAST
35 analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in

Example IV. A chimeric protein is generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both are used as probes to search for homologous 5 genomic sequences from the public human genome databases. Partial DNA sequences are therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences are examined to determine whether they contain a complete gene.

VI. Chromosomal Mapping of KPP Encoding Polynucleotides

The sequences used to assemble SEQ ID NO:57-112 are compared with sequences from the 10 Incyte LIFESEQ database and public domain databases using BLAST and other implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:57-112 are assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research 15 (WIGR), and Généthon are used to determine if any of the clustered sequences have been previously mapped. Inclusion of a mapped sequence in a cluster results in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p- 20 arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and 25 other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a 30 gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST are used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte). This analysis is much faster than 35 multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be

modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \min\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding KPP are analyzed with respect to the tissue sources from which they are derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding KPP. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ database (Incyte, Palo Alto CA).

35 **VIII. Extension of KPP Encoding Polynucleotides**

Full length polynucleotides are produced by extension of an appropriate fragment of the full

length molecule using oligonucleotide primers designed from this fragment. One primer is synthesized to initiate 5' extension of the known fragment, and the other primer is synthesized to initiate 3' extension of the known fragment. The initial primers are designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

Selected human cDNA libraries are used to extend the sequence. If more than one extension is necessary or desired, additional or nested sets of primers are designed.

High fidelity amplification is obtained by PCR using methods well known in the art. PCR is performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contains DNA template, 200 nmol of each primer, reaction buffer containing Mg²⁺, (NH₄)₂SO₄, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ are as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well is determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate is scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture is analyzed by electrophoresis on a 1 % agarose gel to determine which reactions are successful in extending the sequence.

The extended nucleotides are desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides are separated on low concentration (0.6 to 0.8%) agarose gels, fragments are excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells are selected on antibiotic-containing media, and individual colonies are picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid

media.

The cells are lysed, and DNA is amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94 °C, 3 min; Step 2: 94 °C, 15 sec; Step 3: 60 °C, 1 min; Step 4: 72 °C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72 °C, 5 min; Step 7: storage at 4 °C. DNA is quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries are reamplified using the same conditions as described above. Samples are diluted with 20% dimethylsulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in KPP Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) are identified in SEQ ID NO:57-112 using the LIFESEQ database (Incyte). Sequences from the same gene are clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters is used to distinguish SNPs from other sequence variants. Preliminary filters remove the majority of basecall errors by requiring a minimum Phred quality score of 15, and remove sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis is applied to the original chromatogram files in the vicinity of the putative SNP. Clone error filters use statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters use statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removes duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs are selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprises 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezuelan, and two Amish individuals. The African population comprises 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprises 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprises 126 individuals (64 male, 62 female) with a reported parental

breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies are first analyzed in the Caucasian population; in some cases those SNPs which show no allelic variance in this population are not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

5 Hybridization probes derived from SEQ ID NO:57-112 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μ Ci of
10 [γ -³²P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing
15 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II,
15 Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to NYTRAN PLUS nylon membranes (Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium
20 dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*),
25 mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface
30 of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

35 Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof

may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of

5 detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser desorption and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is

10 described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)⁺ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)⁺ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/ μ l oligo-(dT) primer (21mer), 1X 15 first strand buffer, 0.03 units/ μ l RNase inhibitor, 500 μ M dATP, 500 μ M dGTP, 500 μ M dTTP, 40 μ M dCTP, 40 μ M dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)⁺ RNA with GEMBRIGHT kits (Incyte). Specific control poly(A)⁺ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction 20 sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (BD 25 Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μ l 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification 30 uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μ g. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope 35 slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR

Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S.

5 Patent No. 5,807,522, incorporated herein by reference. 1 μ l of the array element DNA, at an average concentration of 100 ng/ μ l, is loaded into the open capillary printing element by a high-speed robotic apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene).

Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water.
10 Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μ l of sample mixture consisting of 0.2 μ g each of Cy3 and

15 Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm² coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μ l of 5X SSC in a corner of the chamber. The chamber containing the arrays is
20 incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an

25 Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a
30 resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores.

Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the
35 signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5.

Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a 5 cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different 10 fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a 15 linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each 20 spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be 25 differentially expressed.

Expression

For example, SEQ ID NO:62, SEQ ID NO:93, SEQ ID NO:95, and SEQ ID NO:100 showed differential expression in colon cancer samples, as determined by microarray expression analysis.

Expression of SEQ ID NO:62 was upregulated by at least two-fold in matched normal versus 30 tumorous colon tissues in one out of thirteen donors tested. Expression of SEQ ID NO:93 was upregulated by at least two-fold in matched tumorous versus normal colon tissues in two out of six donors tested and expression of SEQ ID NO:95 was upregulated by at least two-fold in matched tumorous versus normal colon tissues in one out of fourteen donors tested. Expression of SEQ ID NO:100 showed differential expression in colon cancer samples, as determined by microarray 35 expression analysis. Matched normal and tumor samples from various donors with colon cancer

(Huntsman Cancer Institute, Salt Lake City, UT) were compared by competitive hybridization. In 2 out of 6 tumor samples, the expression of SEQ ID NO:100 was increased at least 2-fold, and up to 7-fold, when compared to matched normal tissue from the same donors. Therefore, in various embodiments, SEQ ID NO:62, SEQ ID NO:93, SEQ ID NO:95, and SEQ ID NO:100 can be used for 5 one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In an alternative example, SEQ ID NO:62-63, SEQ ID NO:82-83, SEQ ID NO:89, and SEQ ID NO:97 showed differential expression in lung cancer samples, as determined by microarray expression analysis. Expression of SEQ ID NO:62 was upregulated by at least two-fold in matched 10 normal versus tumorous lung tissues in five out of ten donors tested and downregulated by at least two-fold in matched normal versus tumorous lung tissues in one out of ten donors tested. Different donors could have different primary lesions and the differential regulation of the expression of SEQ ID NO:62 in the tumorous tissues from different donors might be an indicator of this. Further, expression of SEQ ID NO:63 was upregulated by at least two-fold in matched normal versus 15 tumorous lung tissues in one out of six donors tested, as determined by microarray analysis.

In another example, expression of SEQ ID NO:82 was down-regulated in lung tumor tissue versus matched normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:82 was decreased at least two-fold in lung tumor tissue as compared to normal tissue from the same donor in one out of ten donors. In a similar experiment, expression of SEQ ID NO:83 was upregulated 20 in lung tumor tissue versus matched normal lung tissue as determined by microarray analysis. Expression of SEQ ID NO:83 was decreased at least two-fold in lung tumor tissue as compared to normal tissue from the same donor in three out of four donors. Further, expression of SEQ ID NO:89 and SEQ ID NO:97 were found to be downregulated by at least two-fold in matched tumorous versus normal lung tissues in one out of ten donors tested. They were both downregulated 25 in the same donor in independently done experiments. Therefore, in various embodiments, SEQ ID NO:62-63, SEQ ID NO:82-83, SEQ ID NO:89, and/or SEQ ID NO:97 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

In an alternative example, expression of SEQ ID NO:62 was upregulated by at least two-fold 30 in matched normal versus tumorous ovarian tissues in one donor tested in two different experiments, as determined by microarray analysis. Further, expression of SEQ ID NO:111 was differentially expressed in ovarian tumor samples, as determined by microarray expression analysis. Tissue from a normal ovary from a 79 year-old female donor was compared to an ovarian tumor tissue from the same donor (Huntsman Cancer Institute, Salt Lake City, UT). Expression of SEQ ID NO:111 was 35 increased at least 2-fold in the tumor tissue versus the normal tissue. Therefore, in various

embodiments, SEQ ID NO:62 and SEQ ID NO:111 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

In an alternative example, SEQ ID NO:76, SEQ ID NO:84, SEQ ID NO:88, SEQ ID NO:92,

5 SEQ ID NO:94, and SEQ ID NO:96 showed differential expression in breast cancer samples, as determined by microarray expression analysis. For example, expression of SEQ ID NO:76, SEQ ID NO:92, and SEQ ID NO:96 was down-regulated in breast tumor tissue versus matched normal breast tissue as determined by microarray analysis. Expression of SEQ ID NO:76, SEQ ID NO:92 and SEQ ID NO:96 was found to be downregulated by at least two-fold in matched tumorous versus normal

10 breast tissues in one donor tested.

Further, expression of SEQ ID NO:84 was up-regulated in treated MDA-mb-231 cells versus untreated MDA-mb-231 cells as determined by microarray analysis. MDA-mb-231 is a breast tumor cell line isolated from the pleural effusion of a 51-year old female. It forms poorly differentiated adenocarcinoma in nude mice and ALS treated BALB/c mice. It also expresses the Wnt3 oncogene, EGF, and TGF- α . MDA-MB-231 cells were treated with 10 ng/ml TNF- α for 1, 4, 8, 24, 48, and 72 hours. Treated cells were compared to untreated cells kept in culture for the same duration.

Expression of SEQ ID NO:84 was increased at least two-fold in MDA-mb-231 cells treated with TNF- α for more than eight hours.

Further, BT-20 is a breast carcinoma cell line derived *in vitro* from the cells emigrating out 20 of thin slices of the tumor mass isolated from a 74-year-old female. BT-20 cells were treated with α -estradiol for 4, 8, 14, 24, 36, and 48 hours. These treated cells were compared to untreated BT-20 cells kept in culture for the same amount of time. SEQ ID NO:88 was found to be downregulated by at least two-fold in the treated cells after four hours of treatment and continued to be downregulated by at least two-fold until 36 hours of treatment.

25 Further, MDA-mb-231 is a breast tumor cell line isolated from the pleural effusion of a 51-year old female. It forms poorly differentiated adenocarcinoma in nude mice and ALS treated BALB/c mice. It also expresses the Wnt3 oncogene, EGF, and TGF- α . MDA-MB-231 cells were treated with 10 ng/ml TNF- α for 1, 4, 8, 24, 48, and 72 hours. Treated cells were compared to untreated cells kept in culture for the same duration. SEQ ID NO:94 was found to be upregulated by 30 at least two-fold in the treated cells after a minimum of 24 hours of treatment and continued to be upregulated until 72 hours of treatment. Further, MCF7 is a nonmalignant breast adenocarcinoma cell line isolated from the pleural effusion of a 69-year-old female. MCF7 has retained characteristics of the mammary epithelium such as the ability to process estradiol via cytoplasmic estrogen receptors and the capacity to form domes in culture. MCF7 cells were treated with 10 ng/ml IL-1 β for 1, 4, 8, 12, 24, 36, 48, and 72 hours. Treated cells were compared to untreated cells

kept in culture for the same amount of time. SEQ ID NO:94 was found to be upregulated by at least two-fold in treated cells after a minimum of 4 hours of treatment and a maximum of 24 hours of treatment. Therefore, in various embodiments, SEQ ID NO:76, SEQ ID NO:84, SEQ ID NO:88, SEQ ID NO:92, SEQ ID NO:94, and SEQ ID NO:96 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

In another example, expression of SEQ ID NO:84 was up-regulated in treated human aortic endothelial cells (HAEC) versus untreated HAECs as determined by microarray analysis. HAECs were treated with TNF- α for 1, 2, 4, 6, 8, 10, 24, and 48 hours. These TNF- α treated cells were compared to untreated HAECs. Expression of SEQ ID NO:84 was increased at least two-fold in HAECs treated with TNF- α for more than two hours. Therefore, in various embodiments, SEQ ID NO:84 can be used for one or more of the following: i) monitoring treatment of vascular or inflammatory disorders and related diseases and conditions, ii) diagnostic assays for vascular or inflammatory disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for vascular or inflammatory disorders and related diseases and conditions.

In an alternative example, human aortic endothelial cells (HAECs) are primary cells derived from the endothelium of a human aorta. HAECs have been used as an experimental model for investigating the role of the endothelium in human vascular biology *in vitro*. Activation of the vascular endothelium is considered to be a central event in a wide range of both physiological and pathophysiological processes, such as vascular tone regulation, coagulation and thrombosis, atherosclerosis, and inflammation. HAECs were grown to 85% confluence and then treated with TNF- α for 1, 2, 4, 6, 8, 10, 24, and 48 hours. These TNF- α treated cells were compared to untreated HAECs collected at 85% confluence (0 hour). SEQ ID NO:94 was found to be upregulated by at least two-fold in treated cells after a minimum of 4 hours of treatment and until 48 hours of treatment. Further, HUVECs were grown to 85% confluence and then treated with TNF- α for 0.33, 0.66, 1, 4, 8, 24, 48, and 72 hours. These TNF- α treated cells were compared to untreated HUVECs collected at 85% confluence (0 hour). SEQ ID NO:94 was found to be upregulated by at least two-fold in treated cells after a minimum of 8 hours of treatment and until 72 hours of treatment. Therefore, in various embodiments, SEQ ID NO:94 can be used for one or more of the following: i) monitoring treatment of immune disorders and related diseases and conditions, ii) diagnostic assays for immune disorders and related diseases and conditions, and iii) developing therapeutics and/or other treatments for immune disorders and related diseases and conditions.

In an alternative example, SEQ ID NO:100 showed differential expression in activated vascular endothelium, as determined by microarray expression analysis. HUVECs were pretreated with lipofectin alone or with antisense oligonucleotides plus lipofectin. Specifically, the total

treatment over a 72-hour period included two 4 hour incubations at t=0 hour and t=24 hours. These two treated samples were compared to untreated HUVECs kept in culture for 72 hours. The antisense oligonucleotide was an 18mer shown to specifically target the degradation of the WSB-1 mRNA (Incyte gene ID 423565, GenBank g454669) under the transfection conditions employed.

5 Three HUVEC cultures were treated as described above, but in addition they were stimulated for 24 hours at t=48 hours in the presence (or absence) of recombinant Tumor Necrosis Factor alpha (TNF- α) and Interleukin 1 beta (IL-1 β). Each cytokine-activated culture was compared to its corresponding unactivated culture incubated for the last 24 hours in medium alone. In addition, comparisons were made directly between HUVEC cultures activated with cytokines, i.e. activated
10 HUVECs with lipofectin pretreatment or activated HUVECs with lipofectin /antisense pretreatment were compared to activated HUVECs without pretreatment. Cytokine-treated HUVEC cells showed an increase in SEQ ID NO:100 expression of at least 22-fold, when compared to their control cultures. The treatment of HUVEC cells with lipofectin or lipofectin plus the anti-sense oligo had no effect on this stimulation of SEQ ID NO:100 expression by the cytokine treatment. Therefore, in
15 various embodiments, SEQ ID NO:100 can be used for one or more of the following: i) monitoring treatment of diseases or disorders involving activated vascular endothelium, ii) diagnostic assays for diseases or disorders involving activated vascular endothelium, and iii) developing therapeutics and/or other treatments for diseases or disorders involving activated vascular endothelium.

In another example, SEQ ID NO:93 and SEQ ID NO:100 showed differential expression in

20 Tangier disease-derived fibroblasts, as determined by microarray expression analysis. Human fibroblasts were obtained from skin explants from both normal subjects and two patients with homozygous Tangier disease. Cell lines were immortalized by transfection with human papillomavirus 16 genes E6 and E7 and a neomycin resistance selectable marker. In addition, both types of cells were cultured in the presence of cholesterol and compared with the same cell type
25 cultured in the absence of cholesterol. TD derived cells are shown to be deficient in an assay of apoA-I mediated tritiated cholesterol efflux. SEQ ID NO:93 and SEQ ID NO:100 showed increased expression of at least 2-fold in the TD-derived cells, when compared to normal fibroblasts. The same result was seen irrespective of treatment with cholesterol. Therefore, in various embodiments, SEQ ID NO:93 and SEQ ID NO:100 can be used for one or more of the following: i) monitoring
30 treatment of Tangier disease, ii) diagnostic assays for Tangier disease, and iii) developing therapeutics and/or other treatments for Tangier disease.

In another example, SEQ ID NO:111 showed differential expression in liver cells upon treatment with steroids or toxic compounds, as determined by microarray expression analysis. C3A cells were treated with a number of different substances, including beclomethasone,
35 medroxyprogesterone (MAH), budesonide, dexamethasone (Dex), betamethasone, and catechol.

SEQ ID NO:111 expression increased at least 2-fold in treated C3A cells, compared to untreated C3A cells, at various time points and concentrations: beclomethasone - 10 μ M for 3 h; MAH - 1 μ M for 6 h, 10 μ M for 1, 3, or 6 h, and 100 μ M for 1 h; budesonide - 10 μ M for 1 or 3 h and 100 μ M for 1 h; Dex - 100 μ M for 6 h; betamethasone - 1 μ M for 1 or 6 h, 10 μ M for 6 h, and 100 μ M for 1, 3, 5 or 6 h; and catechol - 10 μ M for 3 h. Therefore, in various embodiments, SEQ ID NO:111 can be used for monitoring treatment with steroids or other potentially toxic compounds.

In an alternative example, SEQ ID NO:62 and SEQ ID NO:71 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference 10 sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a 15 common reference sample, differential expression values are directly comparable from one tissue to another.

In an alternative example, SEQ ID NO:62, SEQ ID NO:71, SEQ ID NO:73, SEQ ID NO:76, SEQ ID NO:79-80, SEQ ID NO:87, SEQ ID NO:89, SEQ ID NO:92, SEQ ID NO:95-97 showed tissue-specific expression as determined by microarray analysis. RNA samples isolated from a 20 variety of normal human tissues were compared to a common reference sample. Tissues contributing to the reference sample were selected for their ability to provide a complete distribution of RNA in the human body and include brain (4%), heart (7%), kidney (3%), lung (8%), placenta (46%), small intestine (9%), spleen (3%), stomach (6%), testis (9%), and uterus (5%). The normal tissues assayed were obtained from at least three different donors. RNA from each donor was 25 separately isolated and individually hybridized to the microarray. Since these hybridization experiments were conducted using a common reference sample, differential expression values are directly comparable from one tissue to another. For example, the expression of SEQ ID NO:62 was increased by at least two-fold in lung and spleen tissues as compared to the reference sample. Therefore, SEQ ID NO:62 can be used as a tissue marker for lung and spleen tissues. In an 30 alternative example, the expression of SEQ ID NO:71 was increased by at least two-fold in heart ventricular tissue as compared to the reference sample. Therefore, SEQ ID NO:71 can be used as a tissue marker for heart ventricular tissue. In an alternative example, the expression of SEQ ID NO:76, SEQ ID NO:92, and SEQ ID NO:96 was increased by at least two-fold in testis as compared to the reference sample. Therefore, SEQ ID NO:76, SEQ ID NO:92, and SEQ ID NO:96 can be used 35 as a tissue marker for testis. In an alternate example, the expression of SEQ ID NO:79 and SEQ ID

NO:80 were both increased by at least two-fold in leukocytes as compared to the reference sample. Therefore, SEQ ID NO:79 and SEQ ID NO:80 can be used as tissue markers for leukocytes. The expression of SEQ ID NO:89 and SEQ ID NO:97 was increased by at least two-fold in blood leukocytes as compared to the reference sample. Therefore, SEQ ID NO:89 and SEQ ID NO:97 can
5 be used as tissue markers for blood leukocytes. In yet another example, the expression of SEQ ID NO:73 was increased by at least two-fold in leukocytes, thymus, spleen, and tonsil as compared to the reference sample. Therefore, SEQ ID NO:73 can be used as a tissue marker for leukocytes, thymus, spleen, and tonsil. The expression of SEQ ID NO:87 was increased by at least two-fold in tonsil tissue as compared to the reference sample. Therefore, SEQ ID NO:87 can be used as a tissue
10 marker for tonsil tissue. The expression of SEQ ID NO:95 was increased by at least two-fold in omentum tissue as compared to the reference sample. Therefore, SEQ ID NO:95 can be used as a tissue marker for omentum tissue.

XII. Complementary Polynucleotides

Sequences complementary to the KPP-encoding sequences, or any parts thereof, are used to
15 detect, decrease, or inhibit expression of naturally occurring KPP. Although use of oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of KPP. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent
20 promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the KPP-encoding transcript.

XIII. Expression of KPP

Expression and purification of KPP is achieved using bacterial or virus-based expression systems. For expression of KPP in bacteria, cDNA is subcloned into an appropriate vector
25 containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express KPP upon induction with isopropyl beta-D-
30 thiogalactopyranoside (IPTG). Expression of KPP in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding KPP by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong
35 polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to

infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

5 In most expression systems, KPP is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from KPP at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). Purified KPP obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, XIX, XX, and XXI, where applicable.

10

15

XIV. Functional Assays

KPP function is assessed by expressing the sequences encoding KPP at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 µg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 µg of an additional plasmid containing sequences 20 encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; BD Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and 25 to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in 30 bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as 35

measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; *Flow Cytometry*, Oxford, New York NY).

The influence of KPP on gene expression can be assessed using highly purified populations

- 5 of cells transfected with sequences encoding KPP and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art.
- 10 Expression of mRNA encoding KPP and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of KPP Specific Antibodies

KPP substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g.,

- Harrington, M.G. (1990) *Methods Enzymol.* 182:488-495), or other purification techniques, is used
- 15 to immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the KPP amino acid sequence is analyzed using LASERGENE software

- (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are
- 20 well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using Fmoc chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-KPP activity by, for example, binding the peptide or KPP to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring KPP Using Specific Antibodies

Naturally occurring or recombinant KPP is substantially purified by immunoaffinity

- 30 chromatography using antibodies specific for KPP. An immunoaffinity column is constructed by covalently coupling anti-KPP antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing KPP are passed over the immunoaffinity column, and the column is

- 35 washed under conditions that allow the preferential absorbance of KPP (e.g., high ionic strength

buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/KPP binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and KPP is collected.

XVII. Identification of Molecules Which Interact with KPP

5 KPP, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled KPP, washed, and any wells with labeled KPP complex are assayed. Data obtained using different concentrations of KPP are used to calculate values for the number, affinity, and association of KPP with the candidate
10 molecules.

Alternatively, molecules interacting with KPP are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (BD Clontech).

KPP may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)
15 which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of KPP Activity

Generally, protein kinase activity is measured by quantifying the phosphorylation of a
20 protein substrate by KPP in the presence of [γ - ^{32}P]ATP. KPP is incubated with the protein substrate, ^{32}P -ATP, and an appropriate kinase buffer. The ^{32}P incorporated into the substrate is separated from free ^{32}P -ATP by electrophoresis and the incorporated ^{32}P is counted using a radioisotope counter. The amount of incorporated ^{32}P is proportional to the activity of KPP. A determination of the specific amino acid residue phosphorylated is made by phosphoamino acid analysis of the
25 hydrolyzed protein.

In one alternative, protein kinase activity is measured by quantifying the transfer of gamma phosphate from adenosine triphosphate (ATP) to a serine, threonine or tyrosine residue in a protein substrate. The reaction occurs between a protein kinase sample with a biotinylated peptide substrate and gamma ^{32}P -ATP. Following the reaction, free avidin in solution is added for binding to the
30 biotinylated ^{32}P -peptide product. The binding sample then undergoes a centrifugal ultrafiltration process with a membrane which will retain the product-avidin complex and allow passage of free gamma ^{32}P -ATP. The reservoir of the centrifuged unit containing the ^{32}P -peptide product as retentate is then counted in a scintillation counter. This procedure allows the assay of any type of protein kinase sample, depending on the peptide substrate and kinase reaction buffer selected. This assay is
35 provided in kit form (ASUA, Affinity Ultrafiltration Separation Assay, Transbio Corporation,

Baltimore MD, U.S. Patent No. 5,869,275). Suggested substrates and their respective enzymes include but are not limited to: Histone H1 (Sigma) and p34^{cd2}kinase, Annexin I, Angiotensin (Sigma) and EGF receptor kinase, Annexin II and *src* kinase, ERK1 & ERK2 substrates and MEK, and myelin basic protein and ERK (Pearson, J.D. et al. (1991) Methods Enzymol. 200:62-81).

5 In another alternative, protein kinase activity of KPP is demonstrated in an assay containing KPP, 50 µl of kinase buffer, 1 µg substrate, such as myelin basic protein (MBP) or synthetic peptide substrates, 1 mM DTT, 10 µg ATP, and 0.5 µCi [γ -³²P]ATP. The reaction is incubated at 30°C for 30 minutes and stopped by pipetting onto P81 paper. The unincorporated [γ -³²P]ATP is removed by washing and the incorporated radioactivity is measured using a scintillation counter. Alternatively,
10 the reaction is stopped by heating to 100°C in the presence of SDS loading buffer and resolved on a 12% SDS polyacrylamide gel followed by autoradiography. The amount of incorporated ³²P is proportional to the activity of KPP.

In yet another alternative, adenylate kinase or guanylate kinase activity of KPP may be measured by the incorporation of ³²P from [γ -³²P]ATP into ADP or GDP using a gamma radioisotope counter. KPP, in a kinase buffer, is incubated together with the appropriate nucleotide mono-phosphate substrate (AMP or GMP) and ³²P-labeled ATP as the phosphate donor. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to separate the mono-, di-, and triphosphonucleotide fractions. The diphosphonucleotide fraction is excised and counted. The radioactivity recovered is
20 proportional to the activity of KPP.

In yet another alternative, other assays for KPP include scintillation proximity assays (SPA), scintillation plate technology and filter binding assays. Useful substrates include recombinant proteins tagged with glutathione transferase, or synthetic peptide substrates tagged with biotin. Inhibitors of KPP activity, such as small organic molecules, proteins or peptides, may be identified
25 by such assays.

In another alternative, phosphatase activity of KPP is measured by the hydrolysis of para-nitrophenyl phosphate (PNPP). KPP is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β -mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62). Alternatively,
30 acid phosphatase activity of KPP is demonstrated by incubating KPP-containing extract with 100 µl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 µl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is
35 proportional to the activity of KPP in the assay.

In the alternative, KPP activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM KPP in a final volume of 30 μ l containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% β -mercaptoethanol and 10 μ M substrate, 32 P-labeled on serine/threonine or tyrosine, as appropriate.

5 Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 μ l of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM Na₄P₂O₇, and 2 mM NaH₂PO₄, then centrifuged at 12,000 $\times g$ for 5 min. Acid-soluble 32 Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

XIX. Kinase Binding Assay

10 Binding of KPP to a FLAG-CD44 cyt fusion protein can be determined by incubating KPP with anti-KPP-conjugated immunoaffinity beads followed by incubating portions of the beads (having 10-20 ng of protein) with 0.5 ml of a binding buffer (20 mM Tris-HCL (pH 7.4), 150 mM NaCl, 0.1% bovine serum albumin, and 0.05% Triton X-100) in the presence of 125 I-labeled FLAG-CD44cyt fusion protein (5,000 cpm/ng protein) at 4 °C for 5 hours. Following binding, beads were
15 washed thoroughly in the binding buffer and the bead-bound radioactivity measured in a scintillation counter (Bourguignon, L.Y.W. et al. (2001) J. Biol. Chem. 276:7327-7336). The amount of incorporated 32 P is proportional to the amount of bound KPP.

XX. Identification of KPP Inhibitors

Compounds to be tested are arrayed in the wells of a 384-well plate in varying concentrations along with an appropriate buffer and substrate, as described in the assays in Example XVIII. KPP activity is measured for each well and the ability of each compound to inhibit KPP activity can be determined, as well as the dose-response kinetics. This assay could also be used to identify molecules which enhance KPP activity.

XXI. Identification of KPP Substrates

25 A KPP “substrate-trapping” assay takes advantage of the increased substrate affinity that may be conferred by certain mutations in the PTP signature sequence of protein tyrosine phosphatases. KPP bearing these mutations form a stable complex with their substrate; this complex may be isolated biochemically. Site-directed mutagenesis of invariant residues in the PTP signature sequence in a clone encoding the catalytic domain of KPP is performed using a method standard in
30 the art or a commercial kit, such as the MUTA-GENE kit from BIO-RAD. For expression of KPP mutants in *Escherichia coli*, DNA fragments containing the mutation are exchanged with the corresponding wild-type sequence in an expression vector bearing the sequence encoding KPP or a glutathione S-transferase (GST)-KPP fusion protein. KPP mutants are expressed in *E. coli* and purified by chromatography.

35 The expression vector is transfected into COS1 or 293 cells via calcium phosphate-mediated

transfection with 20 μ g of CsCl-purified DNA per 10-cm dish of cells or 8 μ g per 6-cm dish. Forty-eight hours after transfection, cells are stimulated with 100 ng/ml epidermal growth factor to increase tyrosine phosphorylation in cells, as the tyrosine kinase EGFR is abundant in COS cells. Cells are lysed in 50 mM Tris·HCl, pH 7.5/5 mM EDTA/150 mM NaCl/1% Triton X-100/5 mM 5 iodoacetic acid/10 mM sodium phosphate/10 mM NaF/5 μ g/ml leupeptin/5 μ g/ml aprotinin/1 mM benzamidine (1 ml per 10-cm dish, 0.5 ml per 6-cm dish). KPP is immunoprecipitated from lysates with an appropriate antibody. GST-KPP fusion proteins are precipitated with glutathione-Sepharose, 4 μ g of mAb or 10 μ l of beads respectively per mg of cell lysate. Complexes can be visualized by PAGE or further purified to identify substrate molecules (Flint, A.J. et al. (1997) Proc. Natl. Acad. 10 Sci. USA 94:1680-1685).

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, 15 and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific 20 embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incite Project ID	Polypeptide SEQ ID NO:	Incite Polypeptide ID	Polynucleotide SEQ ID NO:	Incite Polynucleotide ID	Incite Full Length Clones
7521809	1	7521809CD1	57	7521809CB1	95130607CA2
7520259	2	7520259CD1	58	7520259CB1	95110571CA2
7521738	3	7521738CD1	59	7521738CB1	95147830CA2
7522266	4	7522266CD1	60	7522266CB1	95151756CA2
7523011	5	7523011CD1	61	7523011CB1	95164931CA2
7523290	6	7523290CD1	62	7523290CB1	95170294CA2
7523379	7	7523379CD1	63	7523379CB1	90125549CA2, 95161955CA2, 95175488CA2, 95175572CA2
7523387	8	7523387CD1	64	7523387CB1	95178974CA2
7521804	9	7521804CD1	65	7521804CB1	95132316CA2
7521841	10	7521841CD1	66	7521841CB1	2020168CA2, 95151089CA2
7521886	11	7521886CD1	67	7521886CB1	95130564CA2
7521897	12	7521897CD1	68	7521897CB1	95147457CA2
7521995	13	7521995CD1	69	7521995CB1	95150612CA2
7522018	14	7522018CD1	70	7522018CB1	95142092CA2
7523799	15	7523799CD1	71	7523799CB1	
7521743	16	7521743CD1	72	7521743CB1	
7522317	17	7522317CD1	73	7522317CB1	95151602CA2
7522400	18	7522400CD1	74	7522400CB1	95162448CA2
7523524	19	7523524CD1	75	7523524CB1	95180951CA2
7523542	20	7523542CD1	76	7523542CB1	95082240CA2
7523546	21	7523546CD1	77	7523546CB1	95126563CA2
7523552	22	7523552CD1	78	7523552CB1	95097157CA2
7523564	23	7523564CD1	79	7523564CB1	95079959CA2
7523572	24	7523572CD1	80	7523572CB1	95080283CA2
7523586	25	7523586CD1	81	7523586CB1	95116420CA2, 95116444CA2, 95116544CA2
7523617	26	7523617CD1	82	7523617CB1	
7523625	27	7523625CD1	83	7523625CB1	95083285CA2
7523650	28	7523650CD1	84	7523650CB1	95102960CA2
7523665	29	7523665CD1	85	7523665CB1	95115722CA2

Table 1

Incite Project ID	Polypeptide SEQ ID NO:	Incite Polypeptide ID	Polynucleotide SEQ ID NO:	Incite Polynucleotide ID	Incite Full Length Clones
7523672	30	7523672CD1	86	7523672CB1	95128010CA2, 95128110CA2, 95128801CA2
7523687	31	7523687CD1	87	7523687CB1	95130960CA2
7523689	32	7523689CD1	88	7523689CB1	95127388CA2
7523705	33	7523705CD1	89	7523705CB1	95127964CA2
7523706	34	7523706CD1	90	7523706CB1	95128191CA2
7523707	35	7523707CD1	91	7523707CB1	
7523719	36	7523719CD1	92	7523719CB1	95082516CA2
7523720	37	7523720CD1	93	7523720CB1	
7523737	38	7523737CD1	94	7523737CB1	95103868CA2
7523742	39	7523742CD1	95	7523742CB1	95116748CA2
7523743	40	7523743CD1	96	7523743CB1	95082456CA2
7523745	41	7523745CD1	97	7523745CB1	95127980CA2
7523757	42	7523757CD1	98	7523757CB1	95130835CA2, 95130843CA2, 95130859CA2, 95130919CA2, 95130927CA2, 95130935CA2, 95130943CA2, 95130975CA2, 95131027CA2
7523770	43	7523770CD1	99	7523770CB1	95092760CA2
7523919	44	7523919CD1	100	7523919CB1	95122504CA2
7522140	45	7522140CD1	101	7522140CB1	90172571CA2
7522525	46	7522525CD1	102	7522525CB1	
7525355	47	7525355CD1	103	7525355CB1	
7524443	48	7524443CD1	104	7524443CB1	
7524498	49	7524498CD1	105	7524498CB1	90130072CA2
7524957	50	7524957CD1	106	7524957CB1	
7525097	51	7525097CD1	107	7525097CB1	
7525117	52	7525117CD1	108	7525117CB1	95123286CA2
7516593	53	7516593CD1	109	7516593CB1	
7516603	54	7516603CD1	110	7516603CB1	
7525215	55	7525215CD1	111	7525215CB1	95130903CA2
7525356	56	7525356CD1	112	7525356CB1	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	752189CD1	g4191594	1.4E-62	[Homo sapiens] protein serine/threonine phosphatase 4 regulatory subunit 1 Kloeker, S. et al., Purification and identification of a novel subunit of protein serine/threonine phosphatase 4. J. Biol. Chem. 274, 5339-5347 (1999)
	752189CD1	341798 PPP4R1	1.0E-63	[Homo sapiens] [Regulatory subunit;Protein phosphatase;Hydrolase] Protein phosphatase 4 regulatory subunit 1, a subunit of serine-threonine protein phosphatase 4 that associates with the catalytic subunit (PPP4C) and has repetitive sequences at the N-terminus Kloeker, S. et al., Purification and identification of a novel subunit of protein serine/threonine phosphatase 4. J. Biol Chem 274, 5339-47 (1999).
2	7520259CD1	g36621	1.6E-157	[Homo sapiens] serine/threonine protein kinase Meyerson, M. et al., A family of human cdc2-related protein kinases, EMBO J. 11, 2909-2917 (1992)
	7520259CD1	341558 CDK5	1.2E-158	[Homo sapiens] [Protein kinase;Transferase] Cyclin-dependent protein kinase 5, serine-threonine kinase that associates with the regulatory subunit p35 (CDK5R1) and phosphorylates Tau (MAPT) and beta-catenin (CTNNB1); altered regulation may be involved in the pathogenesis of Alzheimer's disease Meyerson, M. et al., A family of human cdc2-related protein kinases., Embo Journal 11, 2909-17 (1992).
	7520259CD1	583769 Cdk5	2.5E-158	[Mus musculus] [Protein kinase;Transferase] [Cell body (soma);Growth cone] Cyclin-dependent protein kinase 5, serine-threonine kinase that associates with the regulatory subunit p35 (Cdk5r) and phosphorylates neuronal proteins, involved in neuronal differentiation, regulation of myogenesis, and adaptive responses to cocaine Ino, H. et al., Expression of CDK5 (PSSALRE kinase), a neural cdc2-related protein kinase, in the mature and developing mouse central and peripheral nervous systems., Brain Res 661, 196-206 (1994).
3	7521738CD1	g409023	2.4E-94	[Rattus norvegicus] protein tyrosine phosphatase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Qian, Z. et al., Temporal and spatial regulation of the expression of BAD2, a MAP kinase phosphatase, during seizure, kindling, and long-term potentiation, <i>Learn. Memory</i> 1, 180-188 (1994)
7521738CD1	340402 DUSP1	9.4E-57		[Homo sapiens] Protein phosphatase;Hydrolase] Dual specificity phosphatase 1, a stress factor-induced, non-membrane spanning protein phosphatase that inactivates members of the mitogen activated protein kinase family, involved in chemotaxis and possibly heat shock and oxidative stress responses
				Berset, T. et al., Notch Inhibition of RAS Signaling Through MAP Kinase Phosphatase 1. During <i>C. elegans</i> Vulval Development, <i>Science</i> 291, 1055-1058. (2001).
7521738CD1	757310 Ptpn16	1.8E-95		[Rattus norvegicus] Protein phosphatase;Hydrolase] Dual specificity phosphatase 1, a stress factor-induced protein phosphatase that inactivates members of the mitogen activated protein kinase family, involved in cell proliferation, stress responses, and possibly apoptosis
				Keyse, S. M. et al., Oxidative stress and heat shock induce a human gene encoding a protein tyrosine phosphatase, <i>Nature</i> 359, 644-7 (1992).
4	7522266CD1	g13432042	2.2E-115	[Homo sapiens] integrin-linked kinase-associated serine/threonine phosphatase 2C
				Leung-Hagesteijn, C. et al., Modulation of integrin signal transduction by ILKAP, a protein phosphatase 2C associating with the integrin-linked kinase, ILK1, <i>EMBO J.</i> 20, 2160-2170 (2001)
7522266CD1	658342 AF095927	1.1E-115		[Rattus norvegicus] Protein phosphatase;Hydrolase] Protein phosphatase 2C delta, serine-threonine protein phosphatase that is inhibited by magnesium, involved in negative regulation of cell cycle progression possibly through dephosphorylation of proteins crucial for cell proliferation
				Tong, Y. et al., Cloning and characterization of a novel mammalian PP2C isozyme., <i>J Biol Chem</i> 273, 35282-90 (1998).

Table 2

Poly peptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7522266CD1	244265 F33A8.6	3.5E-34	[Caenorhabditis elegans] Member of the protein phosphatase 2C family, which are Mg2+-dependent protein serine and threonine phosphatases, has moderate similarity to a region of rat AF095927, which is a serine-threonine protein phosphatase involved in regulation of cell cycle
5	7523011CD1	g2832753	6.2E-40	Chervitz, S. A. et al., Comparison of the complete protein sets of worm and yeast: Orthology and divergence., <i>Science</i> 282, 2022-2028 (1998). [Homo sapiens] phosphorylase kinase gamma subunit Burwinkel, B. et al., Liver glycogenesis due to phosphorylase kinase deficiency: PHKG2 gene structure and mutations associated with cirrhosis, <i>Hum. Mol. Genet.</i> 7, 149-154 (1998)
	7523011CD1	769162 Phkg2	1.4E-37	[Rattus norvegicus] Protein with high similarity to gamma catalytic subunit of phosphorylase kinase, muscle isoform (rat Phkg1), which phosphorylates and thereby activates glycogen phosphorylase and regulates glycogenolysis, contains two protein kinase domains
	7523011CD1	343636 PHKG1	5.9E-19	Calalb, M. B. et al., Molecular cloning and enzymatic analysis of the rat homolog of "PhK-gamma T," an isoform of phosphorylase kinase catalytic subunit., <i>J Biol Chem</i> 267, 1455-63 (1992).
6	7523290CD1	g6691475	4.8E-33	[Homo sapiens] Protein kinase;Transferase] Gamma catalytic subunit of phosphorylase kinase, muscle isoform, phosphorylates and thereby activates glycogen phosphorylase, regulates glycogenolysis Jones, T. A. et al., Localisation of the gene encoding the catalytic gamma subunit of phosphorylase kinase to human chromosome bands 7p12-q21., <i>Biochim Biophys Acta</i> 1048, 24-9 (1990).
				[Homo sapiens] lysophosphatidic acid phosphatase Hiroyama, M. et al., Purification and characterization of a lysophosphatidic acid-specific phosphatase, <i>Biochem. J.</i> 336 (Pt 2), 483-489 (1998)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7523290CD1	475883 LOC51205	3.5E-34	[Homo sapiens] [Other phosphatase; Hydrolase] [Cytoplasmic; Mitochondrial] Lysophosphatidic acid phosphatase, a mitochondrial enzyme that hydrolyzes lysophosphatidic acid to monoacylglycerol and may play a role in regulation of mitochondrial lipid biosynthesis
				Hiroyama, M. et al., Isolation of a cDNA encoding human lysophosphatidic acid phosphatase that is involved in the regulation of mitochondrial lipid biosynthesis., <i>J Biol Chem</i> 274, 29172-80 (1999).
7	7523379CD1	g5139689	3.1E-16	[Homo sapiens] MOK protein kinase
				Miyata, Y. et al., Molecular cloning and characterization of a novel member of the MAP kinase superfamily., <i>Genes Cells</i> 4, 299-309 (1999)
	7523379CD1	569406 RAGE	2.2E-17	[Homo sapiens] [Protein kinase; Transferase] [Cytoplasmic] Renal tumor antigen, a member of the MAP kinase superfamily, a serine threonine protein kinase containing the classic MAP kinase TEY motif, may play a role in signal transduction, may be a target for T-cell-based immunotherapy of renal cell carcinoma
				Miyata, Y. et al., Molecular cloning and characterization of a novel member of the MAP kinase superfamily., <i>Genes Cells</i> 4, 299-309 (1999).
	7523379CD1	586621 Rage	8.0E-17	[Mus musculus] [Protein kinase; Transferase] [Cytoplasmic] Renal tumor antigen, a member of the MAP kinase superfamily, a serine threonine protein kinase, may play a role in signal transduction; human RAGE may be a target for T-cell-based immunotherapy of renal cell carcinoma
				Miyata, Y. et al., Molecular cloning and characterization of a novel member of the MAP kinase superfamily., <i>Genes Cells</i> 4, 299-309 (1999). (supra)
8	7523379CD1	g7209857	3.9E-102	[Homo sapiens] 43-kDa form skeletal muscle and kidney enriched inositol phosphatase
				Ijuin, T. et al., Identification and Characterization of a Novel Inositol Polyphosphate 5-
				Phosphatase, <i>J. Biol. Chem.</i> (2000) In press

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7523387CD1	476427 SKIP	2.0E-128	[Homo sapiens][Other phosphatase;Hydrolyase] [Cytoplasmic] Skeletal muscle and kidney-enriched inositol phosphatase, an inositol polyphosphate 5-phosphatase that may negatively regulate the actin cytoskeleton by hydrolyzing PtdIns 4,5-bisphosphate
				Ijuin, T. et al., Identification and characterization of a novel inositol polyphosphate 5-phosphatase., J Biol Chem 275, 10870-5 (2000).
	7523387CD1	582575 Pps	2.3E-89	[Mus musculus][Other phosphatase;Hydrolyase] Putative phosphatase, has high similarity to human SKIP, which is a skeletal muscle and kidney enriched inositol phosphatase that may negatively regulate the actin cytoskeleton by hydrolyzing PtdIns 4,5-bisphosphate
				Ijuin, T. et al., Identification and characterization of a novel inositol polyphosphate 5-phosphatase., J Biol Chem 275, 10870-5 (2000). (supra)
9	7521804CD1	g189981	4.6E-83	[Homo sapiens] cAMP-dependent protein kinase subunit RI-beta
				Levy, F. O. et al., Molecular cloning, complementary deoxyribonucleic acid structure and predicted full-length amino acid sequence of the hormone-inducible regulatory subunit of 3'-5'-cyclic adenosine monophosphate-dependent protein kinase from human testis, Mol. Endocrinol. 2, 1364-1373 (1988)
	7521804CD1	337182 PRKAR2B	3.4E-84	[Homo sapiens][Regulatory subunit;Protein kinase;Transferase] Type II regulatory beta subunit of cAMP-dependent protein kinase, mediates subcellular localization of the kinase by associating with A kinase anchor proteins, negatively regulates cell growth and induces cell differentiation in a cAMP-dependent manner
				Chen, Q. et al., Organelle-specific targeting of protein kinase AII (PKAII). Molecular and in situ characterization of murine A kinase anchor proteins that recruit regulatory subunits of PKAII to the cytoplasmic surface of mitochondria., J Biol Chem 272, 15247-57 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7521804CD1	422656 Prka2b	2.3E-46	[Mus musculus] [Protein kinase; Regulatory subunit; Transferase] Type II regulatory beta subunit of cAMP-dependent protein kinase, regulates metabolism, lipolysis, and energy balance
				Singh, I. S. et al., Molecular cloning and characterization of the promoter region of the mouse regulatory subunit RI beta of type II cAMP-dependent protein kinase., Biochem Biophys Res Commun 178, 221-6 (1991).
10	7521841CD1	g556651	1.1E-171	[Homo sapiens] PISSLRE
				Brambilla, R. et al., Molecular cloning of PISSLRE, a novel putative member of the cdk family of protein serine/threonine kinases., Oncogene 9, 3037-3041 (1994).
	7521841CD1	758910 CDK10	2.1E-146	[Homo sapiens] [Protein kinase; Transferase] Cyclin dependent kinase 10, member of the CDK family of serine/threonine kinases that interacts with and inhibits the transcription factor ETS2, involved in cell proliferation, may play a role in regulating the G2/M phase of the cell cycle
				Brambilla, R. et al., Molecular cloning of PISSLRE, a novel putative member of the cdk family of protein serine/threonine kinases., Oncogene 9, 3037-41 (1994).
	7521841CD1	618480 CDC2L1	2.6E-84	[Homo sapiens] [Protein kinase; Transferase] [Nuclear; Cytoplasmic] Cell division cycle 2 like 1, member of the p34 (CDC2) superfamily that contains a PSTAIRE box, a protein kinase involved in apoptosis and cell cycle control; mutation of the corresponding gene is associated with non-Hodgkin lymphoma and melanoma
				Eipers, P. G. et al., Structure and expression of the human p58ckk-1 protein kinase chromosomal gene., Genomics 13, 613-21 (1992).
11	7521886CD1	g1871531	1.7E-30	[Homo sapiens] protein-tyrosine-phosphatase
				Kim, Y. W. et al., Characterization of the PEST family protein tyrosine phosphatase BDP1., Oncogene 13, 2275-2279 (1996).
	7521886CD1	569394 PTPN18	1.2E-31	[Homo sapiens] [Protein phosphatase; Hydrolase] Protein tyrosine phosphatase non-receptor type 18, a tyrosine phosphatase containing a PEST motif
				Kim, Y. W. et al., Characterization of the PEST family protein tyrosine phosphatase BDP1., Oncogene 13, 2275-9 (1996).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7521886CD1	332352 Ptp20	6.2E-18	[Rattus norvegicus] [Protein phosphatase;Hydrolase] [Cytoplasmic] Protein tyrosine phosphatase 20, a member of the protein tyrosine phosphatase-PEST family, may play a role in neuronal differentiation
12	7521897CD1	g2695984	2.2E-106	Aoki, N. et al., The novel protein-tyrosine phosphatase PTP20 is a positive regulator of PC12 cell neuronal differentiation., <i>J Biol Chem</i> 271, 29422-6 (1996).
				[Homo sapiens] protein kinase
				Schliebel, K. et al., Abnormal XY interchange between a novel isolated protein kinase gene, PRKY, and its homologue, PRKX, accounts for one third of all (Y+)XX males and (Y-)XY females., <i>Hum. Mol. Genet.</i> 6, 1985-1989 (1997)
	7521897CD1	618510 PRKY	1.6E-107	[Homo sapiens] [Protein kinase;Transferase] Protein kinase Y1-linked, a putative cAMP dependent serine threonine protein kinase that is encoded on the Y chromosome, may be associated with benign prostatic hyperplasia and carcinoma
				Schliebel, K. et al., FISH localization of the human Y-homolog of protein kinase PRKX (PRKY) to Yp11.2 and two pseudogenes to 15q26 and Xq12->q13., <i>Cytogenet Cell Genet</i> 76, 49-52 (1997).
	7521897CD1	341810 PRKX	1.5E-95	[Homo sapiens] [Protein kinase;Transferase] Protein kinase X-linked, a cAMP-dependent protein kinase catalytic subunit, involved in signaling for macrophage and granulocyte development; gene recombination with the Y-linked PRKY is associated with sex reversal of Swyer syndrome
				Klink, A. et al., The human protein kinase gene PKX1 on Xp22.3 displays Xp/Yp homology and is a site of chromosomal instability., <i>Hum Mol Genet</i> 4, 869-78 (1995).
13	7521995CD1	g1147812	1.7E-37	[Homo sapiens] red cell-type low molecular weight acid phosphatase
				Bryson, G. L. et al., Gene structure, sequence, and chromosomal localization of the human red cell-type low-molecular-weight acid phosphotyrosyl phosphatase gene, ACP1, <i>Genomics</i> 30, 133-140 (1995)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7521995CD1	619032 Acp1	4.0E-33	[<i>Rattus norvegicus</i>] Protein with strong similarity to cytoplasmic phosphotyrosyl protein phosphatase (acid phosphatase, human ACP1), which plays a role in receptor mediated signal transduction, member of the low molecular weight phosphotyrosine protein phosphatase family
	7521995CD1	568344 ACP1	3.4E-27	Fiaschi, T. et al., Low molecular weight protein-tyrosine phosphatase is involved in growth inhibition during cell differentiation., <i>J Biol Chem</i> 276, 49156-63. (2001). [<i>Homo sapiens</i>] [Protein phosphatase;Other phosphatase;Hydrolase][Cytoplasmic] Cytoplasmic phosphotyrosyl protein phosphatase (acid phosphatase), a soluble protein tyrosine phosphatase that plays a role in receptor mediated signal transduction, S isoform may play a role in growth retardation
14	7522018CD1	g1770424	6.5E-203	Wu, Y. Y. et al., Sequencing, cloning, and expression of human red cell-type acid phosphatase, a cytoplasmic phosphotyrosyl protein phosphatase., <i>J Biol Chem</i> 267, 10856-65 (1992).
	7522018CD1	342088 GPRK2L	4.7E-204	Sallese, M. et al., G protein-coupled receptor kinase GRK4. Molecular analysis of the four isoforms and ultrastructural localization in spermatozoa and germinal cells., <i>J. Biol. Chem.</i> 272, 10188-10195 (1997) [<i>Homo sapiens</i>] [Protein kinase;Transferase][Cytoplasmic;Plasma membrane] G protein-coupled receptor kinase 4, a protein kinase that regulates desensitization of G protein-coupled receptors by phosphorylating agonist-stimulated receptors
	7522018CD1	608422 Gprk2I	2.2E-158	Premont, R. T. et al., The GRK4 subfamily of G protein-coupled receptor kinases. Alternative splicing, gene organization, and sequence conservation, <i>J Biol Chem</i> 274, 29381-9 (1999). [<i>Mus musculus</i>] [Protein kinase;Transferase] G protein-coupled receptor kinase 4, a putative protein kinase, may desensitize G protein-coupled receptors by phosphorylating activated receptors

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Premont, R. T. et al., The GRK4 subfamily of G protein-coupled receptor kinases. Alternative splicing, gene organization, and sequence conservation, <i>J Biol Chem</i> 274, 29381-9 (1999). (supra)
15	7523799CD1	g19032238	3.8E-220	[Homo sapiens] protein kinase WNK3
				Verissimo, F. et al., WNK kinases, a novel protein kinase subfamily in multi-cellular organisms, <i>Oncogene</i> 20, 5562-5569 (2001)
	7523799CD1	730131 PRKWNK2	0.0	[Homo sapiens] Protein kinase lysine deficient 2, member of the WNK family of serine/threonine kinases; mutations in the corresponding genes of other family members (PRKWNK1 and PRKWNK4) are associated with a heritable form of hypertension
				Xu Be et al., WNK1, a novel mammalian Serine/Threonine protein kinase lacking the catalytic lysine in subdomain II, <i>J Biol Chem</i> 275, 16795-801 (2000).
	7523799CD1	662422 PRKWNK1	1.4E-221	[Homo sapiens] Protein kinase lysine deficient 1, a serine/threonine kinase
16	7521743CD1	g190422	4.9E-143	Xu Be et al., WNK1, a novel mammalian Serine/Threonine protein kinase lacking the catalytic lysine in subdomain II, <i>J Biol Chem</i> 275, 16795-801 (2000). (supra)
				[Homo sapiens] protein phosphatase-2A subunit-alpha
				Mayer, R. E. et al., Structure of the 55-kDa regulatory subunit of protein phosphatase 2A: evidence for a neuronal-specific isoform, <i>Biochemistry</i> 30, 3589-3597 (1991)
17	7522317CD1	g219902	2.2E-18	[Homo sapiens] protein-tyrosine phosphatase
				Adachi, M. et al., Molecular cloning and chromosomal mapping of a human protein-tyrosine phosphatase LC-PTP, <i>Biochem. Biophys. Res. Commun.</i> 186, 1607-1615 (1992)
		774007 PTPN7	3.7E-19	[Homo sapiens] [Protein phosphatase;Hydrolase][Cytoplasmic] Non-receptor protein tyrosine phosphatase type 7, (hematopoietic tyrosine phosphatase, leukocyte phosphatase), an ERK1 (MAPK3) and ERK2 (MAPK1) phosphatase, involved in T cell receptor signal transduction

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Saxena, M. et al., Inhibition of T cell signaling by mitogen-activated protein kinase-targeted hematopoietic tyrosine phosphatase (HepTP), <i>J Biol Chem</i> 274, 11693-700. (1999).
	330060 HepTP	1.3E-11		[Rattus norvegicus] [Protein phosphatase;Hydrolase] [Cytoplasmic] Hematopoietic tyrosine phosphatase (leukocyte phosphatase), may be involved in IgE receptor mediated signal transduction
				Swieter, M. et al., Aggregation of IgE receptors in rat basophilic leukemia 2H3 cells induces tyrosine phosphorylation of the cytosolic protein-tyrosine phosphatase HepTP., <i>J Biol Chem</i> 270, 21902-6 (1995).
18	7522400CD1	g180178	4.7E-138	[Homo sapiens] cdc2-related protein kinase
				Ninomiya-Tsuji, J. et al., Cloning of a human cDNA encoding a CDC2-related kinase by complementation of a budding yeast cdc28 mutation, <i>Proc. Natl. Acad. Sci. U.S.A.</i> 88, 9006-9010 (1991)
	758846 CDK2	3.4E-139		[Homo sapiens] [Protein kinase;Transferase] [Nuclear;Cytoplasmic] Cyclin-dependent protein kinase 2, interacts with cyclins to regulate kinase activity and cell cycle progression, regulates DNA replication, activates histone gene transcription, abnormal regulation is implicated in various cancers
19	7523324CD1	g11023172	1.1E-227	[Homo sapiens] MAP kinase-interacting kinase 2b
				Slentz-Kesler, K. et al., Identification of the human Mnk2 gene (MKNK2) through protein interaction with estrogen receptor beta, <i>Genomics</i> 69, 63-71 (2000)
	611004 GPRK7	1.6E-228		[Homo sapiens] [Protein kinase;Transferase] G protein-coupled receptor kinase 7 (MAP kinase-interacting kinase), a protein kinase that phosphorylates translation initiation factor 4E (EIF4E) and negatively regulates translation
			—	Chen, C. K. et al., Characterization of human GRK7 as a potential core opsin kinase, <i>Mol Vis</i> 7, 305-13. (2001).

Table 2

Polypeptide Seq ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	618828 Gprk7	1.2E-180	[Mus musculus] [Protein kinase;Transferase] G protein-coupled receptor kinase 7 (MAP kinase-interacting kinase), a serine/threonine kinase that is activated by Mapk1, Mapk11 and Mapk14, phosphorylates initiation factor-4E (Eif4e), and likely regulates translation	
20	7523342CD1	g3820592 625869 PPP2R1B	2.2E-259 1.6E-260	Scheper, G. C. et al., The Mitogen-Activated Protein Kinase Signal-Integrating Kinase Mnk2 Is a Eukaryotic Initiation Factor 4E Kinase with High Levels of Basal Activity in Mammalian Cells., <i>Mol Cell Biol</i> 21, 743-754. (2001). [Homo sapiens] protein phosphatase 2 subunit A isoform, beta [Homo sapiens] [Regulatory subunit;Protein phosphatase;Hydrolase] Protein phosphatase 2A regulatory subunit A beta isoform, regulates protein phosphatase 2A, may be involved in regulation of cell proliferation or cell cycle control; gene mutations are associated with lung, breast and colon cancers
	587791 Ppp2r1a		3.4E-226	Tehrani, M. A. et al., Identification of a novel protein phosphatase 2A regulatory subunit highly expressed in muscle., <i>J Biol Chem</i> 271, 5164-70 (1996). [Mus musculus] [Regulatory subunit;Protein phosphatase;Hydrolase] Protein phosphatase 2 (formerly 2A) regulatory subunit A (PR 65) alpha isoform, a regulatory subunit of the complex; mutations in the human PPP2R1A gene are occasionally detected in cancer cells
21	7523546CD1	g4206721	1.0E-291	Nanahoshi, M. et al., Alpha4 protein as a common regulator of type 2A-related serine/threonine protein phosphatases., <i>FEBS Lett</i> 446, 108-12 (1999). [Homo sapiens] HuCds1 kinase
				Brown, A. L. et al., A human Cds1-related kinase that functions downstream of ATM protein in the cellular response to DNA damage, <i>Proc. Natl. Acad. Sci. U.S.A.</i> 96, 3745-3750 (1999)
	428448 CHEK2		7.4E-293	[Homo sapiens] [Protein kinase;Transferase] [Nuclear] CHK2 checkpoint homolog, protein kinase involved in DNA damage response and cell cycle arrest, phosphorylated by ataxia telangiectasia mutated kinase (ATM), phosphorylates p53 (TP53) and mediates BRCA1 function; downregulated in some breast cancers

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Chehab, N. H. et al., Chk2/hCds1 functions as a DNA damage checkpoint in G(1) by stabilizing p53., <i>Genes And Development</i> 14, 278-88. (2000).
22	587579 Rad53	3.5E-208		[Mus musculus][Protein kinase;Transferase] RAD53 homolog (checkpoint kinase 2), a protein kinase that may transmit DNA damage signals from ataxia telangiectasia mutated kinase (Atm), may prevent entry into mitosis via phosphorylation of Cdc25c; human CHEK2 is downregulated in some breast cancers
				Matsuoka, S. et al., Linkage of ATM to cell cycle regulation by the Chk2 protein kinase., <i>Science</i> 282, 1893-7 (1998).
	7523552CD1	g3025677	7.0E-41	[Homo sapiens] p38beta2 MAP kinase
				Enslen, H. et al., Selective activation of p38 mitogen-activated protein (MAP) kinase isoforms by the MAP kinase kinases MKK3 and MKK6. <i>J. Biol. Chem.</i> 273, 1741-1748 (1998)
	789549 MAPK11	5.0E-42		[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase 11, a MAP kinase that is activated by proinflammatory cytokines, stress, MKK6 (MAP2K6), and estradiol, mediates ATF2 dependent gene expression, may promote cardiac muscle cell hypertrophy, may inhibit apoptosis
				Jiang, Y. et al., Characterization of the structure and function of a new mitogen- activated protein kinase (p38beta). <i>J. Biol Chem</i> 271, 17920-6 (1996).
	689064 Mapk11	5.0E-42		[Mus musculus][Protein kinase;Transferase] Mitogen-activated protein kinase 11, a component of signal transduction pathways involved in cell differentiation, immune and stress response, and DNA damage repair, member of the MAP kinase family
23	7523564CD1	g258001	3.1E-173	Bell, L. M. et al., Hyperosmotic stress stimulates promoter activity and regulates cellular utilization of the serum- and glucocorticoid-inducible protein kinase (Sgk) by a p38 MAPK-dependent pathway. <i>J. Biol Chem</i> 275, 25262-72 (2000).
				[Homo sapiens] calcineurin A catalytic subunit; calmodulin-dependent protein phosphatase catalytic subunit; CaM-PP catalytic subunit

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Muramatsu, T. et al., Molecular cloning and chromosomal mapping of the human gene for the testis-specific catalytic subunit of calmodulin-dependent protein phosphatase (calcineurin A), <i>Biochem. Biophys. Res. Commun.</i> 188, 265-271 (1992)
24	7523572CD1	g200466	3.7E-234	[Mus musculus] phosphoprotein phosphatase
				Kincaid, R. L. et al., Cloning and characterization of molecular isoforms of the catalytic subunit of calcineurin using nonisotopic methods, <i>J. Biol. Chem.</i> 265, 11312-11319 (1990)
	342700 PPP3CC		6.1E-227	[Homo sapiens] [Protein phosphatase;Hydrolyase] Protein phosphatase 3 catalytic subunit gamma isoform (calcineurin A gamma), putative catalytic subunit of calmodulin-regulated protein phosphatase (calcineurin), testis specific
				Muramatsu, T. et al., Molecular cloning and chromosomal mapping of the human gene for the testis-specific catalytic subunit of calmodulin-dependent protein phosphatase (calcineurin A), <i>Biochem Biophys Res Commun</i> 188, 265-71 (1992).
	582573 Ppp3cc		2.7E-235	[Mus musculus] [Protein phosphatase;Hydrolyase] Protein phosphatase 3 catalytic subunit gamma isoform (calcineurin A3), putative catalytic subunit of calmodulin-regulated protein phosphatase (calcineurin), may be involved in spermatogenesis and the regulation of sperm motility
				Muramatsu, T. et al., Molecular cloning of a calmodulin-dependent phosphatase from murine testis: identification of a developmentally expressed nonneural isoenzyme [published erratum appears in <i>Proc Natl Acad Sci U S A</i> 1992 May 15;89(10):4779], <i>Proc Natl Acad Sci U S A</i> 89, 529-33 (1992).
25	7523586CD1	g4519629 341746 OSR1	2.4E-47 1.7E-48	[Homo sapiens] oxidative-stress responsive 1 [Homo sapiens] [Protein kinase;Transferase;Other kinase] Oxidative-stress responsive 1, a putative serine threonine kinase, member of the SOK (Ste20/oxidant stress response kinase) family of kinases that are activated by oxidative stress

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Tanari, M. et al., Isolation and characterization of a novel serine threonine kinase gene on chromosome 3p22-21.3., J Hum Genet 44, 116-20 (1999).
	587833 Sk39	3.3E-38		[Mus musculus][Protein kinase;Transferase][Cytoplasmic;Cytoskeletal] Ste-20-related protein kinase, member of the STE20/SPS1 family, a putative serine/threonine kinase, translocates from the cytosol to the cytoskeleton in response to hyperosmotic shock
26	7523617CD1	E2879898	0.0	Tsutsumi, T. et al., Proline- and alanine-rich Ste20-related kinase associates with F-actin and translocates from the cytosol to cytoskeleton upon cellular stresses., J Biol Chem 275, 9157-62. (2000).
				[Homo sapiens] leucine zipper bearing kinase
				Sakuma, H. et al., Molecular cloning and functional expression of a cDNA encoding a new member of mixed lineage protein kinase from human brain, J. Biol. Chem. 272, 28622-28629 (1997)
	340854 MAP3K13	0.0		[Homo sapiens][Protein kinase;Transferase][Unspecified membrane] Protein kinase with double leucine/isoleucine zippers, member of the mixed-lineage kinase family of proteins that have similarity to both serine/threonine and tyrosine kinases, induces phosphorylation of c-Jun (JUN) and activation of JNK-1 (PRKM8)
				Sakuma, H. et al., Molecular cloning and functional expression of a cDNA encoding a new member of mixed lineage protein kinase from human brain., J Biol Chem 272, 28622-9 (1997).
	586679 Map3k12	2.3E-198		[Mus musculus][Protein kinase;Transferase][Golgi;Nuclear;Cytoplasmic] Mitogen activated protein kinase kinase kinase 12, dual leucine zipper-bearing kinase, a member of the mixed lineage protein kinase (MLK), activates the SAPK/JNK and MAPK signaling pathways, may be involved in the regulation of cell growth and apoptosis

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID or PROTEOME ID NO:	Probability Score	Annotation
				Holzman, I. B. et al., Identification, molecular cloning, and characterization of dual leucine zipper bearing kinase. A novel serine/threonine protein kinase that defines a second subfamily of mixed lineage kinases., <i>J Biol Chem</i> 269, 30808-17 (1994).
27	7523625CD1	g1136410	5.7E-249	[Homo sapiens] similar to protein kinase of <i>X. laevis</i> , has putative transmembrane domain in central region
		570006 MELK	4.1E-250	[Homo sapiens][Protein kinase;Transferase] Protein containing two C-terminal kinase associated domain 1 and two protein kinase domains, has low similarity to microtubule-MAP-affinity regulating kinase (rat LOC60328), which is a serine-threonine kinase that influences microtubule stability
		585291 Mek	1.6E-195	[Mus musculus][Protein kinase;Transferase] Protein containing a protein kinase domain and a C-terminal kinase associated domain 1, has low similarity to rat LOC60328, which is a serine-threonine kinase that participates in microtubule stability and the control of cell polarity
				Gil, M. et al., Cloning and expression of a cDNA encoding a novel protein serine/threonine kinase predominantly expressed in hematopoietic cells., <i>Gene</i> 195, 295-301 (1997).
28	7523630CD1	g6012176	0.0	[Homo sapiens] inducible IkappaB kinase
				Shimada, T. et al., IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases., <i>Int Immunol</i> 11, 1357-1362 (1999)
		569978 IKKE	0.0	[Homo sapiens][Protein kinase;Transferase] IKK-related kinase epsilon; inducible IkappaB kinase, protein with high similarity to TBK-1, TANK -binding kinase, which forms a ternary complex with TANK and TRAF2 to activate NFkappaB, contains a kinase domain
				Tojima, Y. et al., NAK is an IkappaB kinase-activating kinase., <i>Nature</i> 404, 778-82. (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	608728 Ikbbk	1.3E-266	[Mus musculus] [Protein kinase;Transferase] Inhibitor of kappaB kinase epsilon (IkappaB kinase-i), protein kinase induced in response to proinflammatory cytokines and lipopolysaccharide, phosphorylates I kappa B alpha (Nfkbia) and activates NF kappa B, may have a role in the immune response	
				Shimada, T. et al., IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases., <i>Int Immunol</i> 11, 1357-62 (1999).
29	7523665CD1	g6470350	8.5E-171	[Homo sapiens] protein kinase
				Kobayashi, T. et al., Characterization of the structure and regulation of two novel isoforms of serum- and glucocorticoid-induced protein kinase, <i>Biochemical J.</i> 344, 189-197 (1999)
	570898 SGK2	6.1E-172	[Homo sapiens] [Protein kinase;Transferase] Serum-glucocorticoid regulated kinase 2, a serine-threonine kinase that may be regulated by 3-phosphoinositide-dependent kinase-1 (PDK1) and by hydrogen peroxide	
				Kobayashi, T. et al., Characterization of the structure and regulation of two novel isoforms of serum- and glucocorticoid-induced protein kinase, <i>Biochem J</i> 344, 189-97 (1999).
	587325 Sgk2	3.0E-163	[Mus musculus] [Protein kinase;Transferase] Serum-glucocorticoid regulated kinase 2, a putative serine-threonine kinase	
				Kobayashi, T. et al., Characterization of the structure and regulation of two novel isoforms of serum- and glucocorticoid-induced protein kinase., <i>Biochem J</i> 344, 189-97 (1999).
30	7523672CD1	g36619	1.6E-137	[Homo sapiens] serine/threonine protein kinase
				Meyerson, M. et al., A family of human cdc2-related protein kinases, <i>EMBO J.</i> 11, 2909-2917 (1992)
	7523672CD1	743606 PCTK1	1.1E-138	[Homo sapiens] [Protein kinase;Transferase] PCTAIRE protein kinase 1, a serine/threonine protein kinase, activity is cell-cycle regulated and highest at the transition from S to G2 phase; gene mutation may be associated with X chromosome-linked heritable disorders

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Olkuda, T. et al. PCTAIRE-1 and PCTAIRE-3, two members of a novel cdc2/CDC28-related protein kinase gene family., <i>Oncogene</i> 7, 2249-58 (1992).
7523672CD1	586809 Pctk1	6.5E-136		[Mus musculus][Protein kinase;Transferase] PCTAIRE protein kinase 1, a serine/threonine protein kinase that interacts with calpastatin 1 light chain (S100a10) and with several 14-3-3 proteins (Ywhah, Ywhaq, and Ywhaz)
				Sladeczek, F. et al., The Cdk-like protein PCTAIRE-1 from mouse brain associates with p11 and 14-3-3 proteins., <i>Mol Gen Genet</i> 254, 571-7 (1997).
31	7523687CD1	g1575563	3.7E-54	[Homo sapiens] hematopoietic progenitor kinase
7523687CD1	428426 MAP4K1	2.6E-55		[Homo sapiens][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase 1 (hematopoietic progenitor kinase 1), activates the JNK/SAPK signaling pathway, involved in T cell receptor signaling and the response to stress
				Hu, M. C. et al., Human HPK1, a novel human hematopoietic progenitor kinase that activates the JNK/SAPK kinase cascade., <i>Genes And Development</i> 10, 2251-64 (1996).
7523687CD1	320930 Map4k1	1.0E-51		[Mus musculus][Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase kinase 1 (hematopoietic progenitor kinase 1), activates the JNK/SAPK signaling pathway, predicted to be involved in the immune response
				Diener, K. et al., Activation of the c-Jun N-terminal kinase pathway by a novel protein kinase related to human germinal center kinase., <i>Proc Natl Acad Sci U S A</i> 94, 9687-92 (1997).
32	7523689CD1	g1813646	4.4E-262	[Homo sapiens] MEK kinase 3
7523689CD1	336390 MAP3K3	3.1E-263		[Homo sapiens][Protein kinase;Transferase] Mitogen activated protein kinase kinase kinase 3, activates the SAPK (MAPK8) and ERK (MAPK3) but not the p38 MAP kinase (MAPK14) pathway and regulates the cell cycle through cyclin D1 (CCND1) expression, selectively upregulated in hepatocellular carcinoma

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID or PROTEOME ID NO:	Probability Score	Annotation
				Ellinger-Ziegelbauer, H. et al., Direct activation of the stress-activated protein kinase (SAPK) and extracellular signal-regulated protein kinase (ERK) pathways by an inducible mitogen-activated protein Kinase/ERK kinase-kinase kinase 3 (MEKK) derivative., J Biol Chem 272, 2668-74 (1997).
7523689CD1	326414 Map3k3	1.7E-253		[Mus musculus] [Protein kinase; Transferase] Mitogen-activated protein kinase kinase kinase 3, activates the JNK (Mapk8) and Erk (Mapk3), but not the p38 MAP kinase (Mapk14) pathway and is required for embryonic angiogenesis; human MAP3K3 is selectively upregulated in hepatocellular carcinoma
				Blank, J. L. et al., Molecular cloning of mitogen-activated protein/ERK kinase kinases (MEKK) 2 and 3. Regulation of sequential phosphorylation pathways involving mitogen-activated protein kinase and c-Jun kinase., J Biol Chem 271, 5361-8 (1996).
33	7523705CD1	g182574	1.3E-199	[Homo sapiens] p55-c-fgr protein
	7523705CD1	342036 FGR	9.1E-201	[Homo sapiens] [Protein kinase; Transferase] Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog, a member of the Src protein tyrosine kinase family, involved in the viral response, integrin signaling, exocytosis and inhibition of apoptosis, induced in B lymphocytes by Epstein-Barr virus
				Tronick, S. R. et al., Isolation and chromosomal localization of the human fgr protooncogene, a distinct member of the tyrosine kinase gene family., Proc Natl Acad Sci U S A 82, 6595-9 (1985).
	7523705CD1	685827 Fgr	3.5E-160	[Rattus norvegicus] [Protein kinase; Transferase] Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog, has strong similarity to human FGR, which is a member of the Src family of protein tyrosine kinases that is induced in B lymphocytes in response to Epstein-Barr virus
34	7523706CD1	g306833	1.2E-217	[Homo sapiens] protein-tyrosine kinase

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID or PROTEOME ID NO:	Probability Score	Annotation
	7523706CD1	3335754 HCK	9.6E-218	[Homo sapiens][Protein kinase;Transferase] Small molecule-binding protein][Lysosome/vacuole;Golgi;Cytoplasmic;Plasma membrane] Hematopoietic cell kinase, Src family tyrosine kinase involved in monocyte signaling, may play a role in neutrophil degranulation, involved with HIV-1 replication and spreading; mouse Hck mediates the development of encephalomyocarditis-induced diabetes
				Ziegler, S. F. et al., Novel protein-tyrosine kinase gene (hck) preferentially expressed in cells of hematopoietic origin., <i>Mol Cell Biol</i> 7, 2276-85 (1987).
	7523706CD1	584957 Hck	1.6E-193	[Mus musculus][Protein kinase;Transferase][Cytoskeletal] Hematopoietic cell kinase, Src family tyrosine kinase that functions in phagocytosis and interferon gamma (Ifng) signaling in macrophages, may modulate ion transport, mediates activation of macrophages during encephalomyocarditis-induced diabetes
				Hamaguchi, I. et al., Analysis of CSK homologous kinase (CHK/HYL) in hematopoiesis by utilizing gene knockout mice., <i>Biochem Biophys Res Commun</i> 224, 172-9 (1996).
35	7523707CD1	g5558099	0.0	[Homo sapiens] protein kinase C-theta
	7523707CD1	343718 PRKCQ	0.0	[Homo sapiens][Protein kinase;Transferase] Protein kinase C theta, involved in T cell activation and protection from apoptosis; may play a role in insulin and multidrug resistance; rat Pkcq may play a role in hyperglycemia, hypertriglyceridemia and insulin resistance
				Chang, J. D. et al., Molecular cloning and expression of a cDNA encoding a novel isoenzyme of protein kinase C (nPKC). A new member of the nPKC family expressed in skeletal muscle, megakaryoblastic cells, and platelets [published erratum appears in <i>J Biol Chem</i> 1994 Dec 9;269(49):31322], <i>J Biol Chem</i> 268, 14208-14 (1993).
	7523707CD1	582493 Prkcq	0.0	[Mus musculus][Protein kinase;Transferase] Protein kinase C theta, may regulate mitosis, preadipocyte differentiation and spermatogenesis; human PRKCQ may have roles in insulin and multidrug resistance and rat Pkcq may have roles in hyperglycemia, hypertriglyceridemia and insulin resistance

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Osada, S. et al., A new member of the protein kinase C family, nPKC theta, predominantly expressed in skeletal muscle., Mol Cell Biol 12, 3930-8 (1992).
36	7523719CD1	g3820592	7.1E-183	[Homo sapiens] protein phosphatase 2 subunit A isoform beta
	7523719CD1	622869 PPP2R1B	5.1E-184	[Homo sapiens][Regulatory subunit;Protein phosphatase;Hydrolase] Protein phosphatase 2A regulatory subunit A beta isoform, regulates protein phosphatase 2A, may be involved in regulation of cell proliferation or cell cycle control; gene mutations are associated with lung, breast and colon cancers
				Tehrani, M. A. et al., Identification of a novel protein phosphatase 2A regulatory subunit highly expressed in muscle., J Biol Chem 271, 5164-70 (1996).
	7523719CD1	387791 Ppp2r1a	3.0E-154	[Mus musculus][Regulatory subunit;Protein phosphatase;Hydrolase] Protein phosphatase 2 (formerly 2A) regulatory subunit A (PR 65) alpha isoform, a regulatory subunit of the complex; mutations in the human PPP2R1A gene are occasionally detected in cancer cells
37	7523720CD1	g4323326	4.7E-280	Nanahoshi, M. et al., Alpha4 protein as a common regulator of type 2A-related serine/threonine protein phosphatases., FEBS Lett 446, 108-12 (1999).
				[Mus musculus] serine/threonine-protein kinase NEK4
				Chen, A. et al., NTM4-related kinases: isolation and characterization of murine nek3 and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3., Gene 234, 127-137 (1999)
	7523720CD1	430068 Nek4	3.4E-281	[Mus musculus][Protein kinase;Transferase] NTMA (never in mitosis gene a)-related expressed kinase 4, member of NTMA kinase family which controls entrance into mitosis, and may play a role as a cell cycle regulator, highly expressed in testis
				Chen, A. et al., NTM4-related kinases: isolation and characterization of murine nek3 and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3., Gene 234, 127-37 (1999).
	7523720CD1	338322 STK2	8.4E-246	[Homo sapiens][Protein kinase;Transferase] Serine threonine kinase 2, most highly expressed in the heart

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Levedakou, E. N. et al., Two novel human serine/threonine kinases with homologies to the cell cycle regulating Xenopus MO15, and NIMA kinases: cloning and characterization of their expression pattern., <i>Oncogene</i> 9, 1977-88 (1994).
38	7523737CD1	g6012176	0.0	[Homo sapiens] inducible IkappaB kinase
	7523737CD1	569978 IKKE	0.0	[Homo sapiens] Protein kinase;Transferase] IKK-related kinase epsilon; inducible IkappaB kinase, protein with high similarity to TBK-1, TANK -binding kinase, which forms a ternary complex with TANK and TRAF2 to activate NFkappaB, contains a kinase domain
				Peters, R. T. et al., IKKepsilon is part of a novel PMA-inducible IkappaB kinase complex., <i>Mol Cell</i> 5, 513-22 (2000).
	7523737CD1	608728 Ikbbk	2.6E-290	[Mus musculus] Protein kinase;Transferase] Inhibitor of kappaB kinase epsilon (IkappaB kinase-i), protein kinase induced in response to proinflammatory cytokines and lipopolysaccharide, phosphorylates I kappa B alpha (Nfkbia) and activates NF kappa B, may have a role in the immune response
				Shimada, T. et al., IKK-i, a novel lipopolysaccharide-inducible kinase that is related to IkappaB kinases., <i>Int Immunol</i> 11, 1357-62 (1999).
39	7523742CD1	g5225378	7.2E-207	[Homo sapiens] IL-1 receptor-associated kinase-M; IRAK-M
	7523742CD1	428414 IRAK-M	5.2E-208	[Homo sapiens] Protein kinase;Transferase;Receptor (signalling)] Interleukin-1 receptor-associated kinase M, a member of the member of the IRAK-Pelle family, plays a role in interleukin-1 and lipopolysaccharide signal transduction
				Wesche, H. et al., IRAK-M is a novel member of the Pelle/interleukin-1 receptor-associated kinase (IRAK) family., <i>J Biol Chem</i> 274, 19403-10 (1999).
	7523742CD1	429472 Iirak	2.2E-49	[Mus musculus] Protein kinase;Transferase] [Cytoplasmic] IL-1 receptor-associated kinase, protein kinase required for IL1a and IL18 signaling, activates NF-kappaB and JUN kinase, involved in inflammation and the response to bacteria
				Trofimova, M. et al., Developmental and tissue-specific expression of mouse pelle-like protein kinase., <i>J Biol Chem</i> 271, 17609-12 (1996).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
40	7523743CD1	g3820592 625869 PPP2R1B	6.0E-241 4.3E-242	[Homo sapiens] protein phosphatase 2 subunit A isoform beta 2A regulatory subunit A beta isoform; [Protein phosphatase; Hydrolase] Protein phosphatase 2A, may be involved in regulation of cell proliferation or cell cycle control; gene mutations are associated with lung, breast and colon cancers Tehrani, M. A. et al. (supra)
	7523743CD1	587791 Ppp2rla	5.5E-210	[Mus musculus] [Regulatory subunit; Protein phosphatase; Hydrolase] Protein phosphatase 2 (formerly 2A) regulatory subunit A (PR 65) alpha isoform, a regulatory subunit of the complex; mutations in the human PPP2R1A gene are occasionally detected in cancer cells
41	7523745CD1	g182574	1.6E-75	Nanahoshi, M. et al. (supra)
	7523745CD1	342036 FGR	1.1E-76	[Homo sapiens] p55-c-fgr protein [Homo sapiens] [Protein kinase; Transferase] Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog, a member of the Src protein tyrosine kinase family, involved in the viral response, integrin signaling, exocytosis and inhibition of apoptosis, induced in B lymphocytes by Epstein-Barr virus Tronick, S. R. et al. (supra)
	7523745CD1	685827 Fgr	1.9E-49	[Rattus norvegicus] [Protein kinase; Transferase] Gardner-Rasheed feline sarcoma viral (v-fgr) oncogene homolog, has strong similarity to human FGR, which is a member of the Src family of protein tyrosine kinases that is induced in B lymphocytes in response to Epstein-Barr virus Tronick, S. R. et al. (supra)
42	7523757CD1	g531820 341144 MAP4K2	3.1E-70 2.2E-71	[Homo sapiens] GC kinase [Homo sapiens] [Protein kinase; Transferase] [Golgi,Cytoplasmic; Plasma membrane] Mitogen-activated protein kinase kinase kinase kinase 2, member of the germinal center kinase subfamily of serine/threonine protein kinases, activates the stress-activated protein kinase pathway, expression is increased in UV-resistant melanoma cells

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Ren, M. et al., In its active form, the GTP-binding protein rab8 interacts with a stress-activated protein kinase, Proc Natl Acad Sci U S A 93, 5151-5 (1996).
	752375CD1	582031 Map4k2	7.5E-71	[Mus musculus] [Protein kinase;Transferase] [Golgi;Cytoplasmic;Plasma membrane] Mitogen-activated protein kinase kinase kinase kinase 2, member of the germinin center kinase subfamily of serine/threonine protein kinases, interacts with Rab8; expression of human MAP4K2 is increased in UV-resistant melanoma cells
				Ren, M. et al. (supra)
43	7523770CD1	g1813646	0.0	[Homo sapiens] MEK kinase 3
	7523770CD1	336390 MAP3K3	0.0	[Homo sapiens] [Protein kinase;Transferase] Mitogen activated protein kinase kinase kinase 3, activates the SAPK (MAPK8) and ERK (MAPK3) but not the p38 MAP kinase (MAPK14) pathway and regulates the cell cycle through cyclin D1 (CCND1) expression, selectively upregulated in hepatocellular carcinoma
				Ellinger-Ziegelbauer, H. et al. (supra)
	7523770CD1	326414 Map3k3	0.0	[Mus musculus] [Protein kinase;Transferase] Mitogen-activated protein kinase kinase kinase 3, activates the JNK (Mapk8) and Erk (Mapk3), but not the p38 MAP kinase (Mapk14) pathway and is required for embryonic angiogenesis; human MAP3K3 is selectively upregulated in hepatocellular carcinoma
				Blank, J. L. et al. (supra)
44	7523919CD1	g4138209	1.3E-264	[Mus musculus] serine/threonine kinase protein MSTK2L, long-form
				Hayashi, K. et al. Activity and substrate specificity of the murine STK2 serine/threonine kinase that is structurally related to the mitotic regulator protein NIMA of Aspergillus nidulans. Biochem. Biophys. Res. Commun. 264:449-456 (1999).
				[Mus musculus] [Protein kinase; Transferase] NIMA (never in mitosis gene a)-related expressed kinase 4, member of NIMA kinase family which controls entrance into mitosis, and may play a role as a cell cycle regulator, highly expressed in testis
				430068 Nek4
				6.4E-265

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Chen, A. et al. NIMA-related kinases: isolation and characterization of murine nek3 and nek4 cDNAs, and chromosomal localization of nek1, nek2 and nek3. <i>Gene</i> 234:127-137 (1999).
	338322 STK2	7.3E-198		[Homo sapiens] [Protein kinase; Transferase] Serine threonine kinase 2, most highly expressed in the heart
45	7522140CD1	g6729348	1.4E-44	[Arabidopsis thaliana] IRE homolog 1
				Oyama, T. et al. The IRE gene encodes a protein kinase homologue and modulates root hair growth in Arabidopsis. <i>Plant J.</i> 30:289-299 (2002).
	731709 FLJ14813	4.8E-125		[Homo sapiens] Protein containing two protein kinase domains and a protein kinase C-terminal domain, has a region of moderate similarity to a region of <i>S. pombe</i> Orb6p, which is required for the maintenance of cell polarity throughout the cell cycle
	609148 Sast	2.5E-35		[Mus musculus] [Protein kinase; Transferase] [Cytoplasmic; Cytoskeletal] Syntrophin-associated serine/threonine kinase, interacts with syntrophins via PDZ domains, associated with microtubules and microtubule-associated proteins and may link the dystrophin (Dmd)/utrophin (Utrn) network with microtubule filaments
				Lumeng, C. et al. Interactions between beta 2-syntrophin and a family of microtubule-associated serine/threonine kinases. <i>Nat. Neurosci.</i> 2:611-617 (1999).
46	7522525CD1	g6978948	2.6E-178	[Homo sapiens] vaccinia related kinase 3
				Nezu, J. Vaccinia related kinase 3 (VRK3). Published Only in DataBase (2000).
	586137 Vrk1	3.6E-55		[Mus musculus] [Protein kinase; Transferase] [Nuclear] Vaccinia related kinase 1, nuclear serine/threonine protein kinase that undergoes autophosphorylation, similar to vaccinia virus B1R kinase, gene expression is high in proliferative tissues such as testis and spleen
				Zelko, I. et al. Molecular cloning and characterization of a novel nuclear protein kinase in mice. <i>Arch. Biochem. Biophys.</i> 352:31-36 (1998).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID or PROTEOME ID NO:	Probability Score	Annotation
		3388834 VRK1	1.4E-53	[Homo sapiens] [Protein kinase; Transferase] Vaccinia related kinase 1, nuclear serine/threonine kinase that phosphorylates acidic and basic protein substrates, may activate p53 and function in signal transduction, may regulate cellular proliferation, has similarity to vaccinia virus B1R kinase
				Nezu, J. et al. Identification of two novel human putative serine/threonine kinases, VRK1 and VRK2, with structural similarity to vaccinia virus B1R kinase. <i>Genomics</i> 45:327-331 (1997).
				Lopez-Borges, S. et al. The human vaccinia-related kinase 1 (VRK1) phosphorylates threonine-18 within the mdm-2 binding site of the p53 tumour suppressor protein. <i>Oncogene</i> 19:3656-3664 (2000).
47	752535CD1	g18694546	0.0	[Homo sapiens] Full length kinase
				Hock, B. et al. A novel member of the epha receptor family. Patent: WO 0208253-A 31-JAN-2002; MERCK PATENT GmbH (DE).
		340478 EPHA7	1.4E-247	[Homo sapiens] [Protein kinase; Transferase; Receptor (signaling)] [Plasma membrane] Ephrin type A receptor 7, an Eph-related receptor tyrosine kinase, interacts with the PDZ domain of the Ras binding protein AF6 (MLLT4), may be involved in cell signaling
				Fox, G.M. et al. cDNA cloning and tissue distribution of five human EPH-like receptor protein-tyrosine kinases. <i>Oncogene</i> 10:897-905 (1995).
				Hock, B. et al. PDZ-domain-mediated interaction of the Eph-related receptor tyrosine kinase EphB3 and the ras-binding protein AF6 depends on the kinase activity of the receptor. <i>Proc. Natl. Acad. Sci. USA</i> 95:9779-9784 (1998).
		790351 EphA7	1.1E-247	[Rattus norvegicus] [Protein kinase; Transferase; Receptor (signaling)] [Plasma membrane] Ephrin type A receptor 7, an Eph-related receptor tyrosine kinase, may be involved in axon guidance
				Janis, L.S. et al. Ephrin-A binding and EphA receptor expression delineate the matrix compartment of the striatum. <i>J. Neurosci.</i> 19:4962-4971 (1999).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Matsunaga, T. et al. Distinct expression patterns of eph receptors and ephrins relate to the structural organization of the adult rat peripheral vestibular system. <i>Eur. J. Neurosci.</i> 12:1599-1616 (2000).
48	7524443CD1	g4322936	0.0	[Homo sapiens] HPK/GCK-like kinase HGK
				Yao, Z. et al. A novel human STE20-related protein kinase, HGK, that specifically activates the c-Jun N-terminal kinase signaling pathway. <i>J. Biol. Chem.</i> 274:2118-2125 (1999).
		340694 Map4k4	0.0	[Homo sapiens] [Protein kinase; Transferase] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, activates the c-Jun N-terminal kinase (MAPK8) signaling pathway, does not activate the ERK or p38 (CSBP1) kinase pathways, may be involved in TNF-alpha (TNF) signaling
				Yao, Z. et al., <i>supra</i> (1999).
		582239 Map4k4	1.0E-263	[Mus musculus] [Protein kinase; Transferase; Receptor (signaling)] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, interacts with Nck, interacts with MEKK1 (Map3k1) and activates the c-Jun N-terminal kinase (Mapk8) signaling pathway; mutants fail to develop somites or a hindgut
				Su, Y.C. et al. NIK is a new Ste20-related kinase that binds NCK and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain. <i>EMBO J.</i> 16:1279-1290 (1997).
				Xue, Y. et al. Mesodermal patterning defect in mice lacking the Ste20 NCK interacting kinase (NIK). <i>Dev. Suppl.</i> 128:1559-1572 (2001).
49	752498CD1	g1872546	0.0	[Mus musculus] NIK
				Su, Y.C. et al. NIK is a new Ste20-related kinase that binds NCK and MEKK1 and activates the SAPK/JNK cascade via a conserved regulatory domain. <i>EMBO J.</i> 16:1279-1290 (1997).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	582239 Map4k4	0.0	[<i>Mus musculus</i>] [Protein kinase; Transferase; Receptor (signaling)] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, interacts with Nck, interacts with MEKK1 (Map3k1) and activates the c-Jun N-terminal kinase (Mapk8) signaling pathway; mutants fail to develop somites or a hindgut	
			Su, Y.C. et al., <i>supra</i> (1997).	
			Xue, Y. et al., <i>supra</i> (2001).	
	340694 MAP4K4	0.0	[<i>Homo sapiens</i>] [Protein kinase; Transferase] Mitogen-activated protein kinase kinase kinase kinase 4, a serine-threonine kinase, activates the c-Jun N-terminal kinase (MAPK8) signaling pathway, does not activate the ERK or p38 (CSBP1) kinase pathways, may be involved in TNF-alpha (TNF) signaling	
			Yao, Z. et al., <i>supra</i> (1999).	
50	7524957CD1	g8101079	0.0	[<i>Homo sapiens</i>] FYVE domain-containing dual specificity protein phosphatase FYVE-DSP2
			Zhao, R. et al. FYVE-DSP2, a FYVE domain-containing dual specificity protein phosphatase that dephosphorylates phosphatidylinositol 3-phosphate. <i>Exp. Cell Res.</i> 265:329-338 (2001).	
	617920 MTMR4	0.0	[<i>Homo sapiens</i>] [Protein phosphatase; Hydrolase] Myotubularin related protein 4, a member of the myotubularin family of dual specificity protein phosphatases, displays phosphatase activity toward phosphatidylinositol 3-phosphate	
			Taylor, G.S. et al. Inaugural Article: Myotubularin, a protein tyrosine phosphatase mutated in myotubular myopathy, dephosphorylates the lipid second messenger, phosphatidylinositol 3-phosphate. <i>Proc. Natl. Acad. Sci. USA</i> 97:8910-8915 (2000).	
	777224 Mtnr4	0.0	[<i>Mus musculus</i>] Protein with high similarity to myotubularin related protein 3 (human MTMR3), which dephosphorylates inositol lipid 3-phosphates and proteins and is involved in membrane lipid metabolism, contains two FYVE zinc finger domains	

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
51	7525097CD1	g1256003	0.0	[Homo sapiens] tyrosine kinase
				Hoehn, G.T. et al. <i>Tnk1: a novel intracellular tyrosine kinase gene isolated from human umbilical cord blood CD34+/Lin-/CD38- stem/progenitor cells. Oncogene 12:903-913 (1996).</i>
	568496 TNK1		0.0	[Homo sapiens] [Protein kinase; Transferase] Tyrosine kinase non-receptor 1, tyrosine phosphorylated, interacts with phospholipase C gamma 1 (PLCG1), may regulate phospholipid signaling pathways during fetal development and in adult cells of the lymphohematopoietic system
				Hoehn, G.T. et al., <i>supra</i> (1996).
				Felschow, D.M. et al. Characterization of the tyrosine kinase Trnk1 and its binding with phospholipase C-gamma 1. <i>Biochem. Biophys. Res. Commun.</i> 273:294-301 (2000).
	710095 Trnk1		7.8E-258	[Mus musculus] Tyrosine kinase non-receptor 1
52	7525117CD1	g3402293	9.8E-53	[Homo sapiens] aurora and IPL1-like midbody-associated protein kinase-1
				Tatsuka, M. et al. Multinuclearity and increased ploidy caused by overexpression of the aurora- and Ipk1-like midbody-associated protein mitotic kinase in human cancer cells. <i>Cancer Res.</i> 58:4811-4816 (1998).
				Katayama, H. et al. Human AIM-1: cDNA cloning and reduced expression during endomitosis in megakaryocyte-lineage cells. <i>Gene 224:1-7 (1998).</i>
	341306 STK12		6.8E-54	[Homo sapiens] [Protein kinase; Transferase] Serine threonine kinase 12, protein predicted to function in cytokinesis, may be involved in ploidy increases in tumor cells, maximally expressed during G2/M phases
				Terada, Y. et al. AIM-1: a mammalian midbody-associated protein required for cytokinesis. <i>EMBO J. 17:667-676 (1998).</i>
				Kawasaki, A. et al. Downregulation of an AIM-1 Kinase Couples with Megakaryocytic Polyploidization of Human Hematopoietic Cells. <i>J. Cell Biol.</i> 152:275-288 (2001).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID or PROTEOME ID NO:	Probability Score	Annotation
		757286[Sk12	5.0E-35	[Rattus norvegicus] [Protein kinase; Transferase] [Nuclear; Centrosome/spindle pole body] Aurora and Ipk1-like midbody-associated protein, a serine/threonine kinase that phosphorylates myosin II regulatory light chain, involved in cytokinesis and localizes to the cleavage furrow of dividing cells
				Murata-Hori, M. et al. Myosin II regulatory light chain as a novel substrate for AIM-1, an aurora/Ipk1p-related kinase from rat. <i>J. Biochem.</i> (Tokyo) 128:903-907. (2000).
53	7516593CD1	g312998	6.6E-292	[Homo sapiens] (X73458) protein kinase
				Golsteyn, R.M. et al. Cell cycle analysis and chromosomal localization of human Plk1, a putative homologue of the mitotic kinases Drosophila polo and Saccharomyces cerevisiae Cdc5. <i>J. Cell. Sci.</i> 107:1509-1517 (1994).
		341782[PLK	1.8E-291	[Homo sapiens] [Protein kinase; Transferase] Polo-like kinase, regulates CDC2-cyclin B through activation of CDC25C phosphatase, plays a role in multiple mitotic checkpoints, maturation of mitotic centrosomes, and cell proliferation; highly expressed in tumor tissues
				Golsteyn, R.M. et al., Cell cycle regulation of the activity and subcellular localization of Plk1, a human protein kinase implicated in mitotic spindle function. <i>J. Cell Biol.</i> 129:1617-1628 (1995).
				Lane, H.A. et al. Antibody microinjection reveals an essential role for human polo-like kinase 1 (Plk1) in the functional maturation of mitotic centrosomes. <i>J. Cell Biol.</i> 135:1701-1713 (1996).
				Smith, M.R. et al. Malignant transformation of mammalian cells initiated by constitutive expression of the polo-like kinase. <i>Biochem. Biophys. Res. Commun.</i> 234:397-405 (1997).
		385563[Plk	1.5E-278	[Mus musculus] [Protein kinase; Transferase] Polo-like kinase, a serine/threonine protein kinase that plays a role in mitotic cell cycle control, apoptosis, activation of the anaphase promoting complex/cyclosome (APC), and cell proliferation; human Plk is highly expressed in tumor tissues

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
				Clay, F. J. et al., Identification and cloning of a protein kinase-encoding mouse gene, PIK, related to the polo gene of Drosophila., Proc Natl Acad Sci U S A 90, 4882-6 (1993).
				Wianny, F. et al. Mouse polo-like kinase 1 associates with the acentriolar spindle poles, meiotic chromosomes and spindle midzone during oocyte maturation. Chromosoma 107:430-439 (1998).
54	7516603CD1	g1196436	4.6E-94	[Homo sapiens] guanylate kinase
				Fitzgibbon, J. et al. Human guanylate kinase (GUK1): cDNA sequence, expression and chromosomal localisation. FEBS Lett. 385:185-188 (1996).
		335704 GUK1	3.2E-95	[Homo sapiens] [Transferase; Other kinase; Small molecule-binding protein] Guanylate kinase 1, catalyzes the conversion of GMP and GDP during GTP synthesis and the cGMP cycle, may function in phototransduction, involved in activation of antiviral drugs, may be a chemotherapy target; gene is downregulated in diffuse astrocytomas
				Brady, W. A. et al., Cloning, characterization, and modeling of mouse and human guanylate kinases., J Biol Chem 271, 16734-40 (1996).
		583145 Guk1	5.5E-84	[Mus musculus] [Transferase; Other kinase] Guanylate kinase 1, catalyzes the conversion of GMP and GDP during GTP synthesis and the cGMP cycle, involved in the activation of antiviral drugs and may be a chemotherapy target; the human GUK1 gene is downregulated in diffuse astrocytomas
				Brady, W. A. et al., <i>supra</i> (1996).
55	7523215CD1	g531820	0.0	[Homo sapiens] GC kinase
				Katz, P. et al. Differential expression of a novel protein kinase in human B lymphocytes. Preferential localization in the germinal center, J. Biol. Chem. 269:16802-16809 (1994).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		341144 MAP4K2	0.0	[Homo sapiens] [Protein kinase; Transferase] [Golgi; Cytoplasmic; Plasma membrane] Mitogen-activated protein kinase kinase kinase kinase 2, member of the germline center kinase subfamily of serine/threonine protein kinases, activates the stress-activated protein kinase pathway, expression is increased in UV-resistant melanoma cells
				Pombo, C.M. et al., Activation of the SAPK pathway by the human STE20 homologue germline centre kinase., <i>Nature</i> 377, 750-4 (1995).
		582031 Map4k2	0.0	[Mus musculus] [Protein kinase; Transferase] [Golgi; Cytoplasmic; Plasma membrane] Mitogen-activated protein kinase kinase kinase kinase 2, member of the germline center kinase subfamily of serine/threonine protein kinases, interacts with Rab8; expression of human MAP4K2 is increased in UV-resistant melanoma cells
56	7525356CD1	g5410312	1.3E-162	Ren, M. et al. In its active form, the GTP-binding protein rab8 interacts with a stress-activated protein kinase. <i>Proc. Natl. Acad. Sci. USA</i> 93:5151-5155 (1996).
				Zhang, Q.H. et al. Cloning and functional analysis of cDNAs with open reading frames for 300 previously undefined genes expressed in CD34+ hematopoietic stem/progenitor cells. <i>Genome Res.</i> 10:1546-1560 (2000).
		665001 Prkaa2	2.3E-203	[Rattus norvegicus] [Protein kinase; Transferase] Catalytic alpha 2 subunit of the 5'-AMP-activated protein kinase, which is a metabolic sensor of AMP levels; involved in fatty acid beta-oxidation, stimulates glucose transport in skeletal muscle during exercise by activation of nitric oxide synthase
				Verhoeven, A.J. et al. The AMP-activated protein kinase gene is highly expressed in rat skeletal muscle. Alternative splicing and tissue distribution of the mRNA. <i>Eur. J. Biochem.</i> 228:236-243 (1995).
				da Silva Xavier, G. et al. Role of AMP-activated protein kinase in the regulation by glucose of islet beta cell gene expression. <i>Proc. Natl. Acad. Sci. USA</i> 97:4023-4028 (2000).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
56, continued				Mu, J. et al. A role for amp-activated protein kinase in contraction- and hypoxia-regulated glucose transport in skeletal muscle. <i>Mol. Cell</i> 7:1085-1094 (2001).
	343708 PRKAA2	1.6E-202		[Homo sapiens] [Protein kinase; Transferase] Catalytic alpha 2 subunit of the 5'-AMP-activated protein kinase, which is a metabolic sensor of AMP levels; involved in fatty acid beta-oxidation, may mediate exercise-induced glucose uptake in skeletal muscle
				Aguan, K. et al. Characterization and chromosomal localization of the human homologue of a rat AMP-activated protein kinase-encoding gene: a major regulator of lipid metabolism in mammals. <i>Gene</i> 149:345-350 (1994).
				Dagher, Z. et al. The effect of AMP-activated protein kinase and its activator AICAR on the metabolism of human umbilical vein endothelial cells. <i>Biochem. Biophys. Res. Commun.</i> 265:112-115 (1999).
				Musi, N. et al. AMP-activated protein kinase (AMPK) is activated in muscle of subjects with type 2 diabetes during exercise. <i>Diabetes</i> 50:921-927 (2001).

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7521809CD1	229	HEAT repeat: G147-L184, G185-A222 PROTEIN M04C9.6A M04C9.6B M04C9.6C SUBUNIT PHOSPHATASE REGULATORY SERINETHREONINE 3110001J10RIK PD044033; K133-K220	HMMER_PFAM BLAST_PRODOM
			Potential Phosphorylation Sites: S46 S58 S59 S117 S117 T94 T139	MOTIFS
			Potential Glycosylation Sites: N79 N99	MOTIFS
2	7520259CD1	314	Serine/Treonine protein kinases, catalytic domain: T26-F308 Protein kinase domain: T26-F308	HMMER_SMART HMMER_PFAM
			Eukaryotic protein kinase IPB000719: H140-L155, Y201-G211	BLIMPS_BLOCKS
			Receptor tyrosine kinase class V IPB001426: R46-T99, L129-K150, P151-V177	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: I124-K163	BLIMPS_BLOCKS
			Protein Kinases signatures and profile: I124-G174	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: F138-I156, E205-G227	BLIMPS_PRINTS
			Phosphorylase kinase family signature PR01049: E300-F311	BLIMPS_PRINTS
			KINASE TRANSFERASE PROTEIN SERINE/TREONINE PROTEIN ATP BINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: Y189-S309	BLAST_PRODOM
			KINASE PROTEIN TRANSFERASE ATP BINDING SERINE/TREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: L27-D108 F128-R171 Y201-F308	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 I49592 6-276: K6-G13 G35-A299	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 Q02399 6-276: K6-G13 G35-A299	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 S51008 6-286: K6-G13 G35-S309	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P34117 6-275: G35-A299	BLAST_DOMO
			Potential Phosphorylation Sites: S94 S298 T99 T203 T243 T280 Y261	MOTIFS
			Potential Glycosylation Sites: N78	MOTIFS
			Protein Kinase St: V144-I156	MOTIFS
3	7521738CD1	198	Dual specificity phosphatase, catalytic domain: Q23-Q142 Dual specificity phosphatase, catalytic domain: Q23-Q142 Rhodanese-like domain: D9-R122	HMMER_SMART HMMER_PFAM HMMER_PFAM

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
		V87-A97	Tyrosine specific protein phosphatase and dual specificity protein phosphatase family IPB000387:	BLIMPS_BLOCKS
			Rhodanese signatures IPB001307: A34-R47, G38-F48	BLIMPS_BLOCKS
			Tyrosine specific protein phosphatases signature and profiles: E70-I126	PROFILESCAN
			Protein tyrosine phosphatase signature PRO00700: V84-A102	BLIMPS_PRINTS
			PROTEIN PHOSPHATASE DUAL SPECIFICITY HYDROLASE NUCLEAR MAP KINASE	BLAST_PRODOME
			PHOSPHATASE1 MPK1 PD004588: L8-L85	
			PROTEIN PHOSPHATASE HYDROLASE DUAL SPECIFICITY MAP KINASE	BLAST_PRODOME
			PHOSPHATASE1 MPK1 PROTEINYROSINE PD007390: V143-C198	
			VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823 A56115 51-336; R30-F86 L85-A145	BLAST_PRODOME
			PHOSPHATASE; TYROSINE; KINASE; MAP; DM08293 P28562 10-137; A10-A102	BLAST_DOMO
			VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823 I38890 29-320; R30-L85 L85-L144	BLAST_DOMO
			VH1-TYPE DUAL SPECIFICITY PHOSPHATASE DM03823 P28562 169-314; L85-P146	BLAST_DOMO
			Potential Phosphorylation Sites: T107	MOTIFS
			Potential Glycosylation Sites: N129	MOTIFS
			Tyr_Phosphatase_1: V87-I99	MOTIFS
4	7522266CD1	314	Serine/threonine phosphatases, family 2C; D39-V310	HMMER_SMART
			Protein phosphatase 2C: D31-V305	HMMER_PFAM
			Protein phosphatase 2C subfamily IPB000222: E106-F124, D137-A146, A161-V200, R204-D217, D240-K252, S301-V310	BLIMPS_BLOCKS
			PROTEIN PHOSPHATASE 2C MAGNESIUM HYDROLASE MANGANESE MULTIGENE	BLAST_PRODOME
			FAMILY PP2C ISOFORM PD001101: E73-R222	
			PROTEIN PHOSPHATASE 2C DM00377 P494441-299; E73-V310	BLAST_DOMO
			PROTEIN PHOSPHATASE 2C DM00377 P49597 108-431; S89-V310	BLAST_DOMO
			PROTEIN PHOSPHATASE 2C DM00377 P49595 224-490; L125-M308	BLAST_DOMO
			PROTEIN PHOSPHATASE 2C DM00377 P25581 31-300; S104-V310	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S13 S36 S55 S82 S130 S170 S214 S227 S265 S279 T81 T108 T116 T120 T185 T237 T255 Y220 Y285	MOTIFS
5	7523011CD1	101	Potential Glycosylation Sites: N87 N304 PHOSPHORYLASE B KINASE GAMMA CATALYTIC CHAIN SKELETAL MUSCLE ISOFORM SUBUNIT PD027403: E8-E69	MOTIFS BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004P15735I25-281: D25-Q83	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004P07934I20-277: D25-Q83	BLAST_DOMO
			Potential Phosphorylation Sites: T46 T74 T78	MOTIFS
6	7523290CD1	168	Protein_Kinase_Atp: I30-K53 signal_cleavage: M1-H25 Signal_Peptide: M8-H25, M1-A32, M8-A32, M8-C23	MOTIFS SPSCAN HMMER
			Histidine acid phosphatases signatures: R45-L65, G79-Y100	BLIMPS_BLOCKS
			Histidine acid phosphatases signatures: A30-K98	PROFILESCAN
			MAMMALIAN ACID PHOSPHATASE DM07177P15309I1-385: V43-FI138	BLAST_DOMO
			MAMMALIAN ACID PHOSPHATASE DM07177P11117I1-383: L50-FI138	BLAST_DOMO
			Potential Phosphorylation Sites: S7 S129 T149	MOTIFS
			Potential Glycosylation Sites: N161	MOTIFS
			Leucine_Zipper: L10-L31	MOTIFS
			His_Acid_Phosphat_1: L50-P64	MOTIFS
7	7523379CD1	44	Potential Phosphorylation Sites: S23 T14	MOTIFS
			Protein_Kinase_Atp: I10-K33	MOTIFS
8	7523387CD1	240	signal_cleavage: M1-A26	SPSCAN
			Inositol polyphosphate phosphatase, catalytic domain PF00783: R55-L64	BLIMPS_PFAM
			PUTATIVE PHOSPHATASE PUTATIVE PHOSPHONOSTIDE 5PHOSPHATASE TYPE II PD131974: L114-I240	BLAST_PRODOM
			Potential Phosphorylation Sites: S2 S3 S7 S15 S88 S143 S150 S180 S210 S227 T60 T197 T198 Y171	MOTIFS
9	7521804CD1	170	RIIalpha, Regulatory subunit portion of type II PKA R-subunit: A7-E44	HMMER_SMART
			Regulatory subunit of type II PKA R-subunit: A7-E44	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			KINASE REGULATOR CAMP DEPENDENT PROTEIN TYPE CHAIN CAMP BINDING PHOSPHORYLATION DUPLICATION MULTIGENE P008215; M1-L153	BLAST_PRODOM
			CAMP DEPENDENT PROTEIN KINASE TYPE II BETA REGULATORY CHAIN CAMP BINDING PHOSPHORYLATION DUPLICATION P0021086; F51-P80	BLAST_PRODOM
			CAMP-DEPENDENT PROTEIN KINASE REGULATOR CHAIN DM01513 P31323 6-123; A7-E125	BLAST_DOMO
			CAMP-DEPENDENT PROTEIN KINASE REGULATOR CHAIN DM01513 P12369 6-121; A7-E125	BLAST_DOMO
			CAMP-DEPENDENT PROTEIN KINASE REGULATOR CHAIN DM01513 P31322 6-123; A7-E125	BLAST_DOMO
			CAMP-DEPENDENT PROTEIN KINASE REGULATOR CHAIN DM01513 P00515 7-105; G8-T48 S83-E125 F109-R142	BLAST_DOMO
			Potential Phosphorylation Sites: S83 S114 T48 T57 T110	MOTIFS
			Serine/Threonine protein kinases, catalytic domain: I11-F286	HMMER SMART
			Protein kinase domain: C8-F286	HMMER PFAM
			Eukaryotic protein kinase IPB00719; H118-L133	BLIMPS BLOCKS
			Protein kinases signatures and profile: Q102-G152	PROFILESCAN
			DCDRK KINASE P0051525; L217-A312	BLAST PRODOM
			PISSLRE mRNA P046981; S283-P323	BLAST PRODOM
			KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATPBINDING II PHOSPHORYLATION CASEIN ALPHA CHAIN P002608; Y167-F286	BLAST PRODOM
			PROTEIN KINASE DOMAIN DM00004 S49330 41-313; I16-A277	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 JX0296 17-289; I16-A277	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P21127 80-354; I16-A277	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 E54024 4423-697; I16-A277	BLAST DOMO
			Potential Phosphorylation Sites: S46 S89 S239 S318 T22 T135 T159 T181 T220 T276	MOTIFS
			Protein_Kinase_St: I122-M134	MOTIFS
			signal cleavage: M1-A35	SPSCAN
10	7521841CD1	323		
11	7521886CD1	73		

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PHOSPHATASE HYDROLASE PROTEIN PTP TYROSINE PROTEIN TYROSINE PTPK1 FETAL LIVER FL1 PD061766: M1-R53	BLAST_PRODOM
12	7521897CD1	237	Potential Phosphorylation Sites: S7 S10 S52 signal_cleavage: M1-S38 Serine/Threonine protein kinases, catalytic domain: C49-M237 Protein kinase domain: C49-M237 Eukaryotic protein kinase IPB000719: H164-L179 Protein kinases signatures and profile: T148-F201 Tyrosine kinase catalytic domain signature PR00109: M126-R139, Y162-L180 KINASE PROTEIN DOMAIN TRANSFERASE. PD00584: L52-G61 KINASE PROTEIN PKX1 TRANSFERASE SERINE/THREONINE PROTEIN ATP BINDING PD026638: M1-D48	MOTIFS SPSCAN HMMER SMART HMMER PFAM BLIMPS BLOCKS PROFILESCAN BLIMPS PRINTS BLIMPS PRODOM BLAST PRODOM
			PROTEIN KINASE DOMAIN DM00004 P51817 51-287: L52-R200 L173-K232 PROTEIN KINASE DOMAIN DM00004 S41099 118-355: T54-D199 F201-D228 PROTEIN KINASE DOMAIN DM00004 P06245 72-308: T54-L197 F201-H230 PROTEIN KINASE DOMAIN DM00004 P16912 195-431: T54-R200	BLAST DOMO BLAST DOMO BLAST DOMO BLAST DOMO
			Potential Phosphorylation Sites: S81 T70 T115 T233 Y44 Protein Kinase_S: I168-L180 signal_cleavage: M1-S20 Low molecular weight phosphatase family: K7-E80 Low molecular weight phosphotyrosine protein: K7-E80	MOTIFS SPSCAN HMMER SMART HMMER PFAM BLIMPS PRINTS BLIMPS PRINTS BLAST PRODOM
13	7521995CD1	80	LMW phosphotyrosine protein phosphatase signature PR00719: Y9-Y26, I52-G68 Mammalian LMW phosphotyrosine protein phosphatase PR00720: R28-W40 HYDROLASE LOW MOLECULAR WEIGHT PROTEIN PHOSPHATASE PROTEINYROSINEPHOSPHATASE PHOSPHOTYROSINE PTPASE ACID PD002132: M1-G49 PROTEIN-TYROSINE-PHOSPHATASE, LOW MOLECULAR WEIGHT DM01997 P24666 1-156: A2-Q77	BLAST DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN-TYROSINE-PHOSPHATASE, LOW MOLECULAR WEIGHT DM01997/B53874 1-156: A2-Q77	BLAST_DOMO
			PROTEIN-TYROSINE-PHOSPHATASE, LOW MOLECULAR WEIGHT DM01997/P40347 1-160: E3-Q77	BLAST_DOMO
			PROTEIN-TYROSINE-PHOSPHATASE, LOW MOLECULAR WEIGHT DM01997/P41893 1-155: V9-Q77	BLAST_DOMO
			Potential Phosphorylation Sites: S48	MOTIFS
			Potential Glycosylation Sites: N35	MOTIFS
14	7522018CD1_424		Regulator of G protein signalling domain: D20-L140	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: F155-L380	HMMER_SMART
			Regulator of G protein signalling domain: D20-L140	HMMER_PFAM
			Protein kinase domain: F155-L380	HMMER_PFAM
			Eukaryotic protein kinase IPB00719: Q272-L287, Y301-G311	BLIMPS_BLOCKS
			GPCR kinase signature PR00717: F136-Q148, R194-S212	BLIMPS_PRINTS
			Regulator of G protein signalling domain proteins PF00615: Y231-I244	BLIMPS_PFAM
			RECEPTOR G PROTEIN COUPLED KINASE GRK4 SPLICE VARIANT G PROTEIN COUPLED ALPHA PD015194: K373-C424	BLAST_PRODOM
			RECEPTOR KINASE G PROTEIN SIGNAL TRANSDUCTION INHIBITOR REGULATOR OF SIGNALING G PD001580: K27-L140	BLAST_PRODOM
			G PROTEIN COUPLED RECEPTOR KINASE 1 PD091998: Q17-Y158	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 P32298 157-401: H157-R291 A292-I363	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P43249 188-432: Y158-D290 A292-I363	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P43250 188-432: Y158-D290 A292-I363	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 Q09537 205-450: Y158-D290 A292-I363	BLAST_DOMO
			Potential Phosphorylation Sites: S23 S72 S107 S348 S374 S415 T44 T154 T177 T195 Y21 Y347	MOTIFS
			Potential Glycosylation Sites: N342 N372	MOTIFS
			Protein_Kinase_Atp: L161-K184	MOTIFS
			Protein_Kinase_St: I276-L288	MOTIFS

Table 3

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN PHOSPHATASE 2A REGULATORY SUBUNIT PR55 DM02681 S55889 13-513: 110-R127 V149-K267	BLAST_DOMO
			Potential Phosphorylation Sites: S32 S67 S113 S125 S167 S187 S194 S246 T118 T151 T208 T230 T250 Y143	MOTIFS
			Potential Glycosylation Sites: N37	MOTIFS
			Protein Phosphatase 2A Regulatory Subunit Pr55 Signatures Pr55_1: E83-N97	MOTIFS
			Protein Phosphatase 2A Regulatory Subunit Pr55 Signatures Pr55_2: N174-D188	MOTIFS
17	7522317CD1	140	signal_cleavage: M1-T23 Signal Peptide: M1-A32	SPSCAN
			PROTEIN-TYROSINE-PHOSPHATASE, NONRECEPTOR TYPE 5 DM02635 P35236 1-80: M1-R41	HMMER
18	7522400CD1	264	Serine/Threonine protein kinases, catalytic domain: F4-F252 Tyrosine kinase, catalytic domain: F4-F252	HMMER SMART
			Protein kinase domain: F4-F252	HMMER SMART
			Eukaryotic protein kinase IPB00719: H119-L134	HMMER PFAM
			Receptor tyrosine kinase class V IPB001426: K24-Y77, L108-K129, P130-V156	BLIMPS BLOCKS
			Protein kinases signatures and profile: K89-G153	BLIMPS BLOCKS
			Phosphorylase kinase family signature PRO1049: K244-V255	PROFILESCAN
			KINASE TRANSFERASE PROTEIN SERINE/THREONINE PROTEIN ATP-BINDING II	BLIMPS PRINTS
			PHOSPHORYLATION CASEIN ALPHA CHAIN PD002608: V163-F252	BLAST PRODOM
			PROTEIN KINASE DOMAIN DM00004 I48156 6-276: K6-V163 Y159-A243	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P23437 6-286: K6-V163 Y159-Q253	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P24941 6-276: K6-V163 Y159-A243	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P48963 6-276: K6-V163 Y159-A243	BLAST DOMO
			Potential Phosphorylation Sites: S120 S173 S198 S242 T39 T164 T187 Y15	MOTIFS
			Protein_Kinase_Atp: 110-K33	MOTIFS
			Protein_Kinase_St: V123-I135	MOTIFS
19	7523524CD1	459	Serine/Threonine protein kinases, catalytic domain: Y128-V413 Protein kinase domain: Y128-V413	HMMER SMART HMMER PFAM

Table 3

SEQ	Incite ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Eukaryotic protein kinase IPB000719: H242-L257, Y317-G327	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III PB001824: P164-K218, E226-S265, G297-P339	BLIMPS_BLOCKS
			Protein kinases signatures and profile: E226-N281	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: F204-H217, F240-C258, C321-R343	BLIMPS_PRINTS
			Phosphorylase kinase family signature PR01049: A405-C416	BLIMPS_PRINTS
			MAP KINASE INTERACTING KINASE PD047281: M93-V127	BLAST_PRODOM
			KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/TREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: Y128-M207 V231-L278 T294-W412	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 A53300 420-665: E132-F340 Y317-A404	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 A57459 417-662: L130-F340 Y317-A404	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P11798 15-261: L135-F340 Y317-A404	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P18653 409-654: E132-F340 Y317-A404	BLAST_DOMO
			Potential Phosphorylation Sites: S15 S43 S121 S236 S288 S315 S366 T151 T456 Y179	MOTIFS
			Protein_Kinase_Atp: L135-K158	MOTIFS
			Protein_Kinase_St: I246-C258	MOTIFS
20	7523542CD1	537	HEAT repeat: G461-L499, G223-L261, A106-L144, G305-I343, E145-L183, S184-M222, G344-L382, G422-C460, R266-L304, G383-F421, T68-R105, R33-H67, D500-A538	HMMER_PFAM
			PHOSPHATASE SUBUNIT PP2A A PROTEIN REGULATORY REPEAT PR65 MULTIGENE FAMILY PD005088: P39-V533, D19-P497, L121-A535	BLAST_PRODOM
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM01111 A36180 228-385: C176-S334	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM01111 A36180 387-541: N335-K490	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM01111 P30153 386-540: N335-K490	BLAST_DOMO
			Potential Phosphorylation Sites: S59 S148 S167 S184 S204 S334 T46 T90 T195 T227 T465	BLAST_DOMO
21	7523546CD1	586	Forkhead associated domain: N155-V218	MOTIFS
			Serine/Threonine protein kinases, catalytic: Y263-L529	HMMER_SMART
			Tyrosine kinase, catalytic domain: Y263-L529	HMMER_SMART
			FHA domain: Y156-S235	HMMER_PFAM

Table 3

SEQ ID	Incyte Polypeptide NO: ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
	-		Protein kinase domain: Y263-L529	HMMER_PFAM BLIMPS_BLOCKS
			Eukaryotic protein kinase IPB000719: H382-L397, Y447-G457	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: D251-S303, T366-C405, L427-P469	BLIMPS_BLOCKS
			Phorbol esters/diacylglycerol binding domain IPB002219: C151-C167, L381-D390, C428-E437	BLIMPS_BLOCKS
			Protein kinases signatures and profile: A310-S422	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: L344-V357, Y380-L398, V451-H473	BLIMPS_PRINTS
			KINASE PROTEIN TRANSFERASE ATPBINDING SERINE/TREONINEPROTEIN PHOSPHORYLATION RECEPTOR TYROSINEPROTEIN PRECURSOR TRANSMEMBRANE PD000001: D64-E348 Y370-S422 I417-E567	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 S557347 21-266; L269-T519	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P08414 44-285; L269-T519	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P22216 200-456; M265-T519	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 Q09170 169-423; K267-T519	BLAST_DOMO
			Potential Phosphorylation Sites: S5 S61 S67 S73 S116 S148 S183 S295 S303 S399 S400 S478 S485 S499 S561 T114 T176 T248 T285 T315 T366 T519 T560 T575 T576 Y340	MOTIFS
			Potential Glycosylation Sites: N48 N209 N228 N546	MOTIFS
			Protein_Kinase_St: T386-L398	MOTIFS
22	7523552CD1 142		PROTEIN KINASE DOMAIN DM000004 P47811 26-298; L27-G85	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 I49066 26-298; L27-G85	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P533778 29-301; L27-G85	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P533779 66-349; L27-L84	BLAST_DOMO
			Potential Phosphorylation Sites: S40 S56 T16 T68	MOTIFS
			Potential Glycosylation Sites: N14	MOTIFS
			Protein_Kinase_Atp: V30-K53	MOTIFS
23	7523564CD1 325		Protein phosphatase 2A homologues, catalytic domain: L52-Q326	HMMER_SMART
			Calcineurin-like phosphoesterase: A79-D281	HMMER_PFAM
			Serine/threonine specific protein phosphatase IPB000934: P80-D117, I123-S167, A185-P231, S253-Y307	BLIMPS_BLOCKS

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Serine/threonine specific protein phosphatases signature: I123-E168	PROFILESCAN
			Serine/threonine phosphatase family signature PR00114: P80-T107, Y109-H136, L142-Y166, E176-V202, L205-S232, N269-K289, S297-N313	BLIMPS_PRINTS
			PROTEIN PHOSPHATASE SERINE/THREONINE HYDROLASE IRON MANGANESE SUBUNIT MULTIGENE FAMILY CATALYTIC PD000252: L46-N313	BLAST_PRODOME
			PHOSPHOPROTEIN PHOSPHATASE DM001331A38193 36-344: K36-K314	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48452 39-348: G35-K314	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48455 36-344: K36-K314	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48456 34-343: G35-K314	BLAST_DOMO
			Potential Phosphorylation Sites: S2 S9 S131 S204 T11 T23 T27 T73 T157 T203 T248 Y171	MOTIFS
			Ser_ Thr_Phasphatase: L143-E148	MOTIFS
			Protein phosphatase 2A homologues, catalytic domain: L52-M343	HMMER SMART
			Calcineurin-like phosphoesterase: A79-D281	HMMER PFAM
			Serine/threonine specific protein phosphatase IPB000934: P80-D117, I123-S167, A185-P231, S253-Y307	BLIMPS_BLOCKS
			Serine/threonine specific protein phosphatases signature: I123-E168	PROFILESCAN
			Serine/threonine phosphatase family signature PR00114: P80-T107, Y109-H136, L142-C166, E176-I202, L205-S232, N269-K289, S297-N313	BLIMPS_PRINTS
			PROTEIN PHOSPHATASE SERINE/THREONINE HYDROLASE IRON MANGANESE SUBUNIT MULTIGENE FAMILY CATALYTIC PD000252: L46-P334	BLAST_PRODOME
			SUBUNIT PROTEIN PHOSPHATASE CALCINEURIN SERINE/THREONINE HYDROLASE 2B CATALYTIC IRON MANGANESE PD003520: H335-K356 E355-G483	BLAST_PRODOME
			SIMILAR TO SERINE/THREONINE PROTEIN PHOSPHATASE PD112269: H45-P137	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331A38193 36-344: K36-V345	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48452 39-348: G35-V345	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48455 36-344: K36-V345	BLAST_DOMO
			PHOSPHOPROTEIN PHOSPHATASE DM001331P48456 34-343: G35-V345	BLAST_DOMO
24	7523572CD1	488	Potential Phosphorylation Sites: S2 S9 S131 S204 S439 S459 T11 T23 T27 T73 T157 T203 T248 T360 T392 Y171	MOTIFS

Table 3

SEQ ID	Incyte Polypeptide NO: ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
25	7523586CD1	113	Ser_Thr_Phosphatase: L143-E148 signal_cleavage: M1-C36	MOTIFS SPSCAN
			PROTEIN KINASE DOMAIN DM000004 P35465 271-510; E21-G97	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 I49376 270-509; E21-G97	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 Q03497 622-861; I23-G97	BLAST_DOMO
			Potential Phosphorylation Sites: S56 T55	MOTIFS
			Protein_Kinase_Atp: I23-K46	MOTIFS
26	7523617CD1	902	Serine/Threonine protein kinases, catalytic domain: I168-L351 Tyrosine kinase, catalytic domain: I168-L343	HMMER SMART
			Protein kinase domain: I168-I345	HMMER SMART
			Receptor tyrosine kinase class V PB001426: A218-S244, S248-R280, L306-P354	HMMER_PFAM
			Receptor tyrosine kinase class III PB001824: E191-K230, K246-F288, P288-T339	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: G251-I261, V270-V292, C314-F336	BLIMPS_BLOCKS
			LEUCINE ZIPPER BEARING KINASE PD 35277: N550-W902	BLIMPS_PRINTS
			KINASE PROTEIN ZIPPER LEUCINE SERINE/THREONINE SERINE/THREONINE PROTEIN BEARING DLK MUK PD 012433: V84-L174	BLAST_PRODOM
			KINASE PROTEIN ZIPPER LEUCINE SERINE/THREONINE SERINE/THREONINE PROTEIN F33E2.2 BEARING DLK MUK PD 013991: I345-S585 A614-A641	BLAST_PRODOM
			LEUCINE ZIPPER BEARING KINASE PD 062077: M1-S83	BLAST_PRODOM
			KINASE; THREONINE; SERINE; DM05413 A55318 I431-701; K377-S585	BLAST_DOMO
			KINASE; THREONINE; SERINE; DM05413 C2363 398-668; K377-D633	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 A55318 I59-389; E170-T224 L340-I404 H207-F336	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 C2363 126-356; E170-T224 L340-I404 H207-F336	BLAST_DOMO
			Potential Phosphorylation Sites: S39 S71 S111 S241 S244 S267 S335 S361 S508 S515 S534 S630 S647 S709 S747 S751 S752 S761 S790 S792 S805 S837 S843 S877 S882 T224 T236 T313 T353 T450 T482 T890 T893 Y866	MOTIFS
			Potential Glycosylation Sites: N201 N788 N798	MOTIFS
27	7523625CD1	458	Serine/Threonine protein kinases, catalytic domain: Y11-I215	HMMER SMART
			Protein kinase domain: Y11-I215	HMMER_PFAM

Table 3

SEQ	Incyte ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
	Polypeptide ID			
			KINASE PROTEIN KIAA0175 PK38 MATERNAL EMBRYONIC LEUCINE ZIPPER SERINE/THREONINE P69EG3 PD017644; I215-R454	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 S52244 I5-255: L13-E87 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P06782 I57-296: E15-D93 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 F54645 I17-258: L13-E87 E88-M206	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 S51025 I18-258: L13-E87 E88-M206	BLAST_DOMO
			Potential Phosphorylation Sites: S140 S205 S308 S315 T56 T252 T313 T339 T380 T439 T441 T456 Y10 Y379	MOTIFS
			Potential Glycosylation Sites: N306 N437	MOTIFS
			Leucine_Zipper: L117-L138	MOTIFS
			Protein_Kinase_Atp: I17-K40	MOTIFS
28	7523650CD1	597	Serine/Threonine protein kinases, catalytic domain: W9-L293 Protein kinase domain: W9-F294 Protein kinases signatures and profile: E111-E168 Tyrosine kinase catalytic domain signature PR00109: M86-E99, H125-R143, V202-F224 PROTEIN KINASE DOMAIN DM000004 A57459 I61-302: L14-F221 PROTEIN KINASE DOMAIN DM000004 I4910 I17-289: L15-I222 PROTEIN KINASE DOMAIN DM000004 P222216 I200-456: D12-F221 PROTEIN KINASE DOMAIN DM000004 S56639 I53-391: L14-K241 Potential Phosphorylation Sites: S96 T239 T354 T393 T414 T441 Y235 Potential Glycosylation Sites: N41 N452 N520 Protein_Kinase_Atp: L15-K38	HMMER_SMART HMMER_PFAM PROFILE_ESCAN BLIMPS_PRINTS BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO BLAST_DOMO MOTIFS MOTIFS
29	7523665CD1	330	Serine/Threonine protein kinases, catalytic: F35-F292 Protein kinase domain: F35-F292 Eukaryotic protein kinase IPB000719: H151-L166, Y211-G221 Receptor tyrosine kinase class V IPB001426: A140-K161, P162-E188, T193-Y225 Receptor tyrosine kinase class III IPB001824: K73-R127, R135-V174, T191-P233 Protein kinases signatures and profile: R135-P187 Tyrosine kinase catalytic domain signature PR00109: L113-Q126, Y149-L167, V215-Q237	HMMER_SMART HMMER_PFAM BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS PROFILE_ESCAN BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			KINASE PROTEIN TRANSFERASE ATP-BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001:R100-E183 T190-Q252 A261-F292	BLAST_PRODOM
			PROTEIN KINASE C ALPHA DM04692 A37237 1-676: L22-D34 N29-G329	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P12688 349-588: L38-G278	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P54644 122-362: L38-I277	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q06226 100-339: F37-G278	BLAST_DOMO
			Potential Phosphorylation Sites: S279 T33 T306 Y61	MOTIFS
			Protein_Kinase_Atp: I41-K64	MOTIFS
			Protein_Kinase_St: I155-L167	MOTIFS
30	7523672CD1	335	KINASE SERINE/THREONINE PROTEIN TRANSFERASE ATPBINDING PCTAIRE1 PROTEIN PCTAIRE2 PCTAIRE3 CRK5 ALTERNATIVE PD007333: P67-T164	BLAST_PRODOM
			SERINE/THREONINE PROTEIN KINASE PCTAIRE1 TRANSFERASE ATPBINDING CRK5 ALTERNATIVE SPLICING PD020837: M1-A66	BLAST_PRODOM
			KINASE; THREONINE; ATP; SERINE; DM04841 Q00536 53-164: P53-Y165	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q00536 166-436: I166-N289	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q00537 193-463: I166-V265	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q04899 122-392: I166-Y265	BLAST_DOMO
		{	Potential Phosphorylation Sites: S58 S65 S78 S89 S110 S119 S146 S153 S286 T14 T88 T225 T232 T283 Y176	MOTIFS
			Atp_Gtp_A: A177-S184	MOTIFS
			Protein_Kinase_Atp:L171-K194	MOTIFS
31	7523687CD1	122	PROTEIN KINASE DOMAIN DM00004 A53714 17-262: L19-G107	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P27636 27-262: L19-L108	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P08458 20-262: L23-E114	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 I49592 6-276: L20-C94	BLAST_DOMO
			Potential Phosphorylation Sites: S98 T27 T68 T112	MOTIFS
			Protein_Kinase_Atp: I23-K46	MOTIFS
32	7523689CD1	532	signal_cleavage: M1-C63	SPSCAN

Table 3

SEQ D NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Serine/Threonine protein kinases, catalytic domain: W393-C532	HMMER_SMART BLAST_PRODOM
			MAPK/ERK KINASE MEK MEKK TRANSFERASE SERINE/THREONINE PROTEIN ATP BINDING PD024779: M1-Q42 D75-W393	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 A48084 98-248: R395-A524	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P53349 405-658: R395-A524	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q01389 1176-1430: R395-K522	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q10407 826-1084: R395-K522	BLAST_DOMO BLAST_DOMO
			Potential Phosphorylation Sites: S39 S67 S143 S146 S178 S231 S249 S270 S281 S290 S342 S376	MOTIFS
			S438 S481 T30 T348 T415 T433 T497 T501 Y510	MOTIFS
			Potential Glycosylation Sites: N46 N159	HMMER_SMART HMMER_SMART
33	7523705CD1	410	Src homology 2 domains: E142-G232	HMMER_SMART HMMER_PFAM
			Src homology 3 domains: T80-D137	HMMER_SMART HMMER_PFAM
			SH2 domain: W144-Y226	BLIMPS_BLOCKS
			SH3 domain: T80-V136	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: L251-F303	BLIMPS_PRINTS
			Receptor tyrosine kinase class II IPB002011: Q308-Q352	BLIMPS_PRINTS
			Tyrosine kinase catalytic domain signature PR00109: T334-K347	BLIMPS_PRINTS
			SH2 domain signature PR00401: W144-L158, Q165-T175, K177-D188, H194-D204, Q215-V229	BLIMPS_PRINTS
			SH3 domain signature PR00452: T80-A90, D94-N109, E111-L120, K124-V136	BLIMPS_PRINTS
			FGR CFGFR PROTOONCOGENE TYROSINE PROTEIN KINASE P55 FGR TRANSFERASE ATP BINDING PHOSPHORYLATION SH2 PD014967: M1-V79	BLAST_PRODOM
			SH2 DOMAIN KINASE SH3 PROTEIN PHOSPHORYLATION TYROSINE PROTEIN TRANSFERASE ATP BINDING TYROSINE PD000093: W144-C241	BLAST_PRODOM
			SRC HOMOLOGY 2 (SH2) DOMAIN DM00048 P05441 60-283: D137-S261	BLAST_DOMO BLAST_DOMO
			SRC HOMOLOGY 2 (SH2) DOMAIN DM00048 P09769 136-243: V136-M244	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P99769 264-504: T264-Y366	BLAST_DOMO BLAST_DOMO
			SRC HOMOLOGY 2 (SH2) DOMAIN DM00048 P09324 148-272: D137-R260	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S122 S183 S218 S285 S299 S341 S380 S401 T14 T70 T92 T99 T110 T175 T190 T211 T292 Y208	MOTIFS
			Potential Glycosylation Sites: N52 N55 N108 N283	MOTIFS
			Protein_Kinase_Atp: L269-K291	MOTIFS
34	7523706CD1	436	Src homology 2 domains: D63-G142	HMMER SMART
			Serine/Threonine protein kinases, catalytic domain: L172-T426	HMMER SMART
			Tyrosine kinase, catalytic domain: L172-L421	HMMER SMART
			SH2 domain: G68-Y136	HMMER PFAM
			Protein kinase domain: L172-L421	HMMER PFAM
			Receptor tyrosine kinase class V IPB001426: G142-A180, M188-I241, S272-R293, A294-E320, G326-M358, E359-G383, Y384-Y432	BLIMPS BLOCKS
			Receptor tyrosine kinase class III IPB001824: P25-P34, K267-K306, R324-I366, I366-I417, W160-F212	BLIMPS BLOCKS
			Receptor tyrosine kinase class II IPB002011: N217-S261, Y287-A338, N373-I417	BLIMPS BLOCKS
			Protein kinases signatures and profile: K267-N319	PROFILESCAN
			Receptor tyrosine kinase class II signature: N296-G342	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: T243-K256, F281-Y299, F329-I339, S348-G370, C392-F414	BLIMPS PRINTS
			SH2 domain signature PR00401: R87-D98, T104-D114, T125-G139	BLIMPS PRINTS
			TYROSINE PROTEIN KINASE HCK HEMOPOIETIC CELL TRANSFERASE	BLAST PRODOM
			PHOSPHORYLATION ATP BINDING MYRISTYLATION SH2 PD019152: M1-E54	BLAST PRODOM
			PROTEIN KINASE DOMAIN DM00004 P08631 261-503: S171-F414	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P25911 245-488: S171-F414	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 Q07014 245-488: S171-F414	BLAST DOMO
			PROTEIN KINASE DOMAIN DM00004 P07948 245-488: S171-F414	BLAST DOMO
			Potential Phosphorylation Sites: S17 S93 S121 S153 S171 S250 S372 S419 T37 T72 T104 T128 T201 T322 T334 T345 T426 Y321 Y384	MOTIFS
			Protein_Kinase_Tyr: Y287-V299	MOTIFS
35	7523707CD1	643	Protein kinase C conserved region 1 (C1): H232-C281, H160-C209	HMMER SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
35, continued			Extension to Ser/Thr-type protein kinases: R572-N635	HMMER_SMART
			Serine/Threonine protein kinases, catalytic domain: F380-M634	HMMER_SMART
			Phorbol esters/diacylglycerol binding domain: H232-C281, H160-C209	HMMER_PFAM
			Protein kinase domain: F380-F628	HMMER_PFAM
			Protein kinase C terminal domain: R572-M638	HMMER_PFAM
			Eukaryotic protein kinase IPB000719: H496-L511	BLIMPS_BLOCKS
			Phorbol esters/diacylglycerol binding domain IPB002219: H160-C176, Q186-C201, C540-E549	BLIMPS_BLOCKS
			Phorbol esters / diacylglycerol binding domain: F244-Q309	PROFILESCAN
			Phorbol esters / diacylglycerol binding domain: Q169-R233	PROFILESCAN
			Protein kinases signatures and profile: R480-E528	PROFILESCAN
			Diacylglycerol/phorbol-ester binding signature PR00008: V157-T171, C245-G254, Q258-V269, H270-G282	BLIMPS_PRINTS
			PROTEIN KINASE C THETA TYPE NPKC THETA ATP BINDING TRANSFERASE SERINE/THREONINE PROTEIN PHORBOL ESTER PD031783; G282-D379	BLAST_PRODOM
			KINASE PROTEIN PHORBOL ESTER BINDING SERINE/THREONINE PROTEIN TRANSFERASE ATP BINDING ZINC C TYPE PD007344: M52-F122-L123-K158	BLAST_PRODOM
			PROTEIN KINASE C TYPE ATP BINDING TRANSFERASE SERINE/THREONINE PROTEIN PHORBOL ESTER BINDING ZINC PD012824; M1-T51	BLAST_PRODOM
			KINASE PROTEIN SERINE/THREONINE PROTEIN TRANSFERASE ATP BINDING PHORBOL ESTER BINDING C ZINC DUPLICATION PD000351: R572-P636	BLAST_PRODOM
			PROTEIN KINASE C ALPHA DM04692 P05773 L-672, R145-G282 K374-P555 I546-N635 N26-F60	BLAST_DOMO
			PROTEIN KINASE C ALPHA DM04692 A37237 1-676: R145-N284 P360-R558 I546-N635	BLAST_DOMO
			PROTEIN KINASE C ALPHA DM04692 P05130 1-638: R145-G282 E363-P555 R553-K593	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q04759 381-622: I381-V552 E549-G560	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Potential Phosphorylation Sites: S10 S58 S91 S323 S348 S594 S599 S613 T51 T85 T243 T425	MOTIFS
			Potential Glycosylation Sites: N630	MOTIFS
		Dag_Pe_Binding_Domain: H160-C209, H232-C281	MOTIFS	
		Protein_Kinase_Atp:L386-K409	MOTIFS	
36	7523719CD1	556	HEAT repeat: G480-L518, G287-L325, A170-L208, E209-L247, S248-M286, G363-L401, G94-H131, G441-C479, G402-F440, G55-V92, T132-R169, G18-L54, D519-A556, R330-I362 PHOSPHATASE SUBUNIT PP2A A PROTEIN REGULATORY REPEAT PR65 MULTIGENE FAMILY PD005088: D19-L493 I25-V552 D188-A554	HMMER_PFAM
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P54612 1-225; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P30153 1-225; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 A36180 2-226; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P36179 3-227; D19-A239	BLAST_DOMO
			Potential Phosphorylation Sites: S43 S48 S123 S212 S231 S248 S268 S353 T59 T70 T110 T154 T259 T291 T484	MOTIFS
37	7523720CD1	728	Serine/Threonine protein kinases, catalytic domain: Y6-T261 Tyrosine kinase, catalytic domain: Y6-P259	HMMER_SMART
			Protein kinase domain: Y6-I261	HMMER_SMART
			Eukaryotic protein kinase IPB000719: H123-F138	HMMER_PFAM
			Receptor tyrosine kinase class V IPB001426: R26-L79, F112-K133, T134-C160	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001874: S43-E97, Q107-K146, A163-A205	BLIMPS_BLOCKS
			Protein kinases signatures and profile: Q107-H159	PROFILESCAN
			Tyrosine kinase catalytic domain signature PR00109: M83-K96, C121-L139, S187-K209, Y230-V252	BLIMPS_PRINTS
			PROTEIN LSTK1 KINASE LIKE SERINE/THREONINE PROTEIN KINASE NRK2 SERINE/THREONINE TRANSFERASE ATP BINDING MITOSIS PD037124: I265-P453	BLAST_PRODOM
			PROTEIN LSTK1 KINASE LIKE SERINE/THREONINE PROTEIN KINASE NRK2 SERINE/THREONINE TRANSFERASE ATP BINDING MITOSIS PD017660: K563-A726	BLAST_PRODOM

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			SERINE/THREONINE PROTEIN KINASE NRK2 EC 2.7.1. SERINE/THREONINE 2 TRANSFERASE ATP BINDING MITOSIS NUCLEAR PROTEIN PHOSPHORYLATION PD126493; D457-A513	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 P51957 8-251; Y8-Y252	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P51954 6-248; L9-Y252	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P51955 10-261; L9-V252	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 Q08942 22-269; V11-Y252	BLAST_DOMO
			Potential Phosphorylation Sites: S16 S43 S179 S243 S251 S301 S343 S377 S404 S480 S553 S561 S568 S578 S588 S593 S615 S631 S640 S645 S677 T67 T201 T276 T348 T536 T642 T669 Y32	MOTIFS
			Potential Glycosylation Sites: N41 N355	MOTIFS
			Protein Kinase Atp: V12-K35	MOTIFS
38	7523737CD1	646	Serine/Threonine protein kinases, catalytic domain: W9-Y311	HMMER_SMART
			Protein kinase domain: W9-F299	HMMER_PFAM
			Eukaryotic protein kinase IPB00719; R127-M142	BLIMPS_BLOCKS
			Protein kinases signatures and profile: E111-E168	PROFILESCAN
			PROTEIN KINASE DOMAIN DM000004 I4910 17-289; L15-I222	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 P22216 200-456; D12-F221	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 S56639 153-391; L14-K241	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 A57459 61-302; L14-F221	BLAST_DOMO
			Potential Phosphorylation Sites: S96 S612, T239 T414 T453 T474 T501 Y235	MOTIFS
			Potential Glycosylation Sites: N41 N512 N580	MOTIFS
39	7523742CD1	385	Signal Peptide: M1-G26	HMMER
			Serine/Threonine protein kinases, catalytic domain: F104-N380	HMMER_SMART
			Tyrosine kinase, catalytic domain: F104-L382	HMMER_PFAM
			Protein kinase domain: F104-S384	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB001824: T270-V312	BLIMPS_BLOCKS
			Receptor tyrosine kinase class II IPB002011: V154-T198	BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: Y180-Q193, T294-D316, F353-M375	BLIMPS_PRINTS

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM00004 P51617 214-483; K106-L335	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 A49332 43-302; H105-M375	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 Q05652 215-487; D107-M375	BLAST_DOMO
			PROTEIN KINASE DOMAIN DM00004 P43293 70-342; L126-M375	BLAST_DOMO
			Potential Phosphorylation Sites: S49 S51 S54 S150 S258 S273 S291 S340 S374 T171 T173 MOTIFS	MOTIFS
40	7523743CD1	469	HEAT repeat: G287-L325, A170-L208, G369-I407, E209-I247, S248-M286, G94-H131, R330-L368, G408-L446, G55-V92, T132-R169, G18-L54	HMMER_PFAM
			PHOSPHATASE SUBUNIT PP2A A PROTEIN REGULATORY REPEAT PR65 MULTIGENE FAMILY Y PD005088: D19-H466 Y139-L455 E31-I407 I25-L455	BLAST_PRODOM
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P54612 1-225; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P30153 1-225; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 A36180 2-226; A14-A239	BLAST_DOMO
			PHOSPHATASE; TRANSFORMING; 61K; PDF1; DM02202 P36179 3-227; D19-A239	BLAST_DOMO
			Potential Phosphorylation Sites: S43 S48 S123 S212 S231 S248 S268 S398 T39 T70 T110 T154 T259 T291	MOTIFS
41	7523745CD1	147	Src homology 3 domains: T80-D137	HMMER_SMART
			SH3 domain: T80-V136	HMMER_PFAM
			SH3 domain signature PR00452: T80-A90, D94-N109, E111-L120, K124-V136	BLIMPS_PRINTS
			FGR CFGR PROTOONCOGENE TYROSINE PROTEIN KINASE P55FGR BINDING PHOSPHORYLATION SH2 DOMAIN SH3 PROTEIN DOMAIN SH2 KINASE PHOSPHORYLATION ATPBINDING	BLAST_PRODOM
			TYROSINEPROTEIN TRANSFERASE PROTOONCOGENE PD000066; V79-A134	BLAST_DOMO
			SRC HOMOLOGY 3 (SH3) DOMAIN DM00025 P09769 78-134; G78-P135	BLAST_DOMO
			SRC HOMOLOGY 3 (SH3) DOMAIN DM00025 Q02977 81-137; G78-P135	BLAST_DOMO
			SRC HOMOLOGY 3 (SH3) DOMAIN DM00025 B491 485-141; G78-P135	BLAST_DOMO
			SRC HOMOLOGY 3 (SH3) DOMAIN DM00025 A25389 83-139; G78-P135	BLAST_DOMO
			Potential Phosphorylation Sites: S122 T14 T70 T92 T99 T110 MOTIFS	MOTIFS
42	7523757CD1	145	Serine/Threonine protein kinases, catalytic domain: F15-S145	HMMER_SMART
			PROTEIN KINASE DOMAIN DM00004 A53714 17-262; L17-K137	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM000004 P08458 20-262; V21-K137 PROTEIN KINASE DOMAIN DM000004 P55465 271-510; R14-S145	BLAST_DOMO BLAST_DOMO
			PROTEIN KINASE DOMAIN DM000004 I49376 270-509; L18-S145 Potential Phosphorylation Sites: S7 S96 T25	BLAST_DOMO MOTIFS
43	7523770CD1	653	signal_cleavage: M1-C63 Serine/Threonine protein kinases, catalytic domain: W389-A649 Tyrosine kinase, catalytic domain: W389-F648 Protein kinase domain: W389-A649	SPSCAN HMMER_SMART HMMER_SMART HMMER_PFAM
			Eukaryotic protein kinase IPB000719: H508-L523, Y571-G581 Receptor tyrosine kinase class V IPB001426: T497-K518, G519-C545, S553-V585 Receptor tyrosine kinase class III IPB001824: N377-T429, T429-K483, V492-K531, M551-P593 Protein kinases signatures and profile: V492-T543	BLIMPS_BLOCKS BLIMPS_BLOCKS BLIMPS_BLOCKS
			Tyrosine kinase catalytic domain signature PR00109: M470-K483, Y506-R524, G556-I566, A575-Y597 MAPK/ERK KINASE MEK MEKK TRANSFERASE SERINE/THREONINE PROTEIN ATP BINDING PD024779; M1-Q42 D75-W389	PROFILESCAN BLIMPS_PRINTS BLAST_PRODOM
			KINASE PROTEIN TRANSFERASE ATP BINDING SERINE/THREONINE PROTEIN PHOSPHORYLATION RECEPTOR TYROSINE PROTEIN PRECURSOR TRANSMEMBRANE PD000001: N388-I500 A487-L541 F535-W594 Q609-H646 PROTEIN KINASE DOMAIN DM000004 A48084 98-348; R391-A640 PROTEIN KINASE DOMAIN DM000004 P53349 405-658; R391-F638	BLAST_PRODOM BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000004 Q01389 1176-1430; R391-A640 PROTEIN KINASE DOMAIN DM000004 Q10407 826-1084; R391-A640 Potential Phosphorylation Sites: S39 S67 S143 S146 S178 S231 S249 S266 S277 S286 S338 S372	BLAST_DOMO BLAST_DOMO MOTIFS
44	7523919CD1	706	Protein kinase domain: Y6-I172 Serine/Threonine protein kinases, catalytic domain: Y6-I172 Eukaryotic protein kinase IPB000719: H34-F49, Y94-G104	HMMER_PFAM HMMER_SMART BLIMPS_BLOCKS

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Protein kinases signatures and profile: G18-C71	PROFILESCAN
			KINASE SERINE/THREONINE PROTEIN ATP-BINDING TRANSFERASE TYROSINE PROTEIN MITOsis PHOSPHORYLATION NEK4 KINASE-LIKE PD017660: K474-F706 PD037124: I176-P364	BLAST_PRODOM
			KINASE SERINE/THREONINE PROTEIN ATP-BINDING TRANSFERASE TYROSINE PROTEIN MITOsis PHOSPHORYLATION MSTK2L LONG-FORM NEK4 PD126493: D368-A424	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 P51957 8-251: Y8-K36, Q31-V163 DM00004 P48479 9-280:L9-G29, I38-P161 DM00004 P51955 10-261:L9-K30, H25-V163 DM00004 P51954 6-248: Q31-V163	BLAST_DOMO
			Potential Phosphorylation Sites: S16, S90, S154, S162, S212, S254, S288, S315, S391, S464, S472, S479, S489, S499, S504, S526, S542, S551, S556, S588, S637, S691, T112, T187, T259, T447, T553, T580, Y32	MOTIFS
			Potential Glycosylation Sites: N266	MOTIFS
			Serine/Threonine protein kinases active-site signature: I38-L50	MOTIFS
45	7522140CD1	243	Protein kinase domain: F35-V243 Serine/Threonine protein kinases, catalytic domain: F35-I224 Eukaryotic protein kinase IPB000719: H148-L163	HMMER_PFAM HMMER_SMART BLIMPS_BLOCKS
			Protein kinases signatures and profile: M132-G185	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			PROTEIN KINASE DOMAIN DM00004 A56155 714-1002; V38-L177 DM00004 P38679 238-527; V38-S178 DM00004 P43565 796-1240; I37-S229 DM00004 P53894 353-658; V38-S178	BLAST_DOMO
			Potential Phosphorylation Sites: S7, S31, S207, S225, S234	MOTIFS
			Potential Glycosylation Sites: N73	MOTIFS
			Serine/Threonine protein kinases active-site signature: I152-I164	MOTIFS
46	752252CD1	416	PROTEIN KINASE DOMAIN DM00004 C56711 45-301; L167-Y388 DM00004 P20505 18-271; F142-Q371 DM00004 P21098 13-255; F142-C360 DM00004 P42158 11-271; I183-Y388	BLAST_DOMO
			Potential Phosphorylation Sites: S48, S54, S59, S71, S78, S83, S90, S92, S104, S143, S210, S215, S221, S278, S290, S310, S405, T101, T381	MOTIFS
			Potential Glycosylation Sites: N53, N248	MOTIFS
47	752355CD1	839	Signal Peptide: M1-A33 signal_cleavage: M1-A33 Ephrin receptor ligand binding domain: E35-C211 SAM domain (Sterile alpha motif): T762-Q826 Fibronectin type III domain: E288-S368 Protein kinase domain: V476-L731 Fibronectin type 3 domain: P285-S368 Sterile alpha motif: S761-R828	HMMER SPSCAN HMMER_PFAM HMMER_PFAM HMMER_PFAM HMMER_SMART HMMER_SMART

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Tyrosine kinase, catalytic domain: V476-L731	HMMER_SMART
			Receptor tyrosine kinase class V IPB001426: E35-E60, R100-E134, K172-A225, P252-Q276, P498-M551, L562-L581, L582-A603, A604-A630, S636-W668, E669-G693, F694-K742, L755-T800	BLIMPS_BLOCKS
			Receptor tyrosine kinase class III IPB00184: A514-R568, Q577-K616, T634-R676, R676-I727	BLIMPS_BLOCKS
			Receptor tyrosine kinase class II IPB002011: T528-Q572, Y597-T648, G683-I727	PROFILESCAN
			Protein kinases signatures and profile: Q577-E629	BLIMPS_PRINTS
			Tyrosine kinase catalytic domain signature PR00109: T554-R567, Y591-V609, S639-L649, S658-D680, C702-F724	
			RECEPTOR KINASE TYROSINE PROTEIN ATP-BINDING TRANSFERASE EPHRIN PRECURSOR PHOSPHORYLATION TRANSMEMBRANE GLYCOPROTEIN PD001495: E35-C211	BLAST_PRODOM
			PD149648: A213-A284	
			RECEPTOR TYROSINE KINASE CLASS V DM00501 I48611 34-382: I37-E283 DM00501 I48612 34-382: I37-E283 DM00501 P54759 33-382: V36-E283 DM00501 S51741 33-382: V36-E283	BLAST_DOMO
			Potential Phosphorylation Sites: S44, S62, S184, S203, S244, S304, S456, S513, S517, S636, S656, S682, S810, T32, T108, T121, T133, T162, T214, T224, T232, T319, T382, T447, T450, T800, Y335, Y597, Y632	MOTIFS
			Potential Glycosylation Sites: N317	MOTIFS
			Cell attachment sequence: R278-D280	MOTIFS
			EGF-like domain signature 2: C269-C282	MOTIFS

Table 3

SEQ ID	Incyte Polypeptide NO: ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
MOTIFS				
48	7524443CD1	1384	CNH domain: Y1066-R1364 Protein kinase domain: F25-T289 Domain found in NIK1-like kinases, mouse citron and yeast ROM1, ROM2: Y1066-R1364 Serine/Threonine protein kinases, catalytic domain: F25-I289 Protein kinases signatures and profile: W129-T181 ATP-BINDING TRANSFERASE NIK KINASE SERINE/THREONINE PROTEIN PD147187_E545-S736, Q622-K996, S847-W1080 KINASE SERINE/THREONINE PROTEIN BINDING PHORBOL ESTER ATP-BINDING TRANSFERASE GDP-GTP EXCHANGE RHO1 CDC42-BINDING PD014445_L1084-S1208, F1239_S1362 KINASE SERINE/THREONINE PROTEIN ATP-BINDING TRANSFERASE MIG-15 TYROSINE PROTEIN 2.7.1. PD147188_I289-Q624, V750-I897, S960-W1080 PROTEIN KINASE DOMAIN DM00004 A53714 I7-262: I27-S279 DM00004 P08458 I20-262: V31-S279 DM00004 P0676 I18-272: I27-P278 DM00004 P38692 I24-266: E29-S279	HMMER_PFAM HMMER_PFAM HMMER_SMART HMMER_SMART PROFILESCAN BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_PRODOM BLAST_DOMO
			Potential Phosphorylation Sites: S9, S17, S77, S112, S255, S259, S264, S324, S326, S579, S598, S630, S643, S648, S665, S676, S689, S691, S712, S716, S735, S787, S788, S795, S847, S848, S872, S886, S892, S921, S929, S1045, S1128, S1188, S1208, S1248, S1261, S1362, T59, T124, T187, T222, T309, T319, T351, T588, T684, T705, T854, T855, T975, T981, T1041, T1161, T1222, Y321, Y323, Y467 Potential Glycosylation Sites: N33, N577, N732, N884, N983, N1316 Protein kinases ATP-binding region signature: V31-K54	MOTIFS MOTIFS

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			MOTIFS	
49	7524498CD1	1230	CNH domain: Y912-R1210 Protein kinase domain: F25-I289 Domain found in NIK1-like kinases, mouse ciron and yeast ROM1, ROM2, Y912-R1210 Serine/Threonine protein kinases, catalytic domain: F25-I289 Protein kinases signatures and profile: W129-T181 KINASE SERINE/THREONINE PROTEIN BINDING PHORBOL ESTER ATP-BINDING TRANSFERASE GDP-GTP EXCHANGE RHO1 CDC42-BINDING PD014445: L930-S1054, F1085-S1208 ATP-BINDING TRANSFERASE NIK KINASE SERINE/THREONINE PROTEIN PD147187: E514-W926 KINASE SERINE/THREONINE PROTEIN ATP-BINDING TRANSFERASE MIG-15 TYROSINE-PROTEIN 2.7.1. PD147188: I289-P500, D766-I807, V893-W926	HMMER_PFAM HMMER_PFAM HMMER_SMART HMMER_SMART PROFILESCAN BLAST_PRODOM BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 A53714 17-262: L27-S279 DM00004 P08458 20-262: V31-S279 DM00004 P10676 18-272: L27-P278 DM00004 P38692 24-266: E29-S279	BLAST_DOMO
			Potential Phosphorylation Sites: S9, S17, S77, S112, S255, S259, S264, S324, S326, S543, S556, S561, S578, S589, S602, S604, S625, S629, S648, S700, S701, S708, S757, S758, S782, S802, S891, S974, S1034, S1054, S1094, S1107, S1208, T52, T59, T124, T187, T222, T309, T319, T351, T597, T618, T764, T765, T887, T1007, T1068, Y321, Y323, Y467 Potential Glycosylation Sites: N33, N645, N794, N1162 Serine/Threonine protein kinases active-site signature: V149-L161	MOTIFS MOTIFS MOTIFS

Table 3

SEQ ID NO:	Incye Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
50	7524957CD1	1199	FYVE zinc finger: T1113-V1179 Protein present in Fab1, YOTB, Vac1, and EEA1: T1110-V1179	HMMER_PFAM
			Tyrosine specific protein phosphatase and dual specificity protein phosphatase family IPB000387: V409-P419	HMMER_SMART
			Tyrosine specific protein phosphatases signature and profiles: M389-E465	BLIMPS_BLOCKS
			Protein tyrosine phosphatase signature PR00700: P406-L424	PROFILESCAN
			HYDROLASE MYOTUBULARIN PHOSPHATASE SPECIFICITY DOMAIN-CONTAINING DUAL FYVE FIS CDNA RELATED PD014611: L91-A262, G274-Y337, P338-H460, N463-N547	BLAST_PRODOM
			KINASE HYDROLASE FYVE CDNA FACTOR CHROMOSOME ZINC-FINGER RHORAC RECEPTOR PD002336: W1115-I1177	BLAST_PRODOM
			Potential Phosphorylation Sites: S94, S213, S216, S238, S299, S375, S379, S602, S614, S620, S653, S661, S686, S692, S698, S705, S722, S740, S773, S849, S856, S909, S938, S996, S1083, S1172, T107, T122, T144, T204, T582, T631, T673, T828, T861, T927, T1110, Y254, Y433, Y1036	MOTIFS
			Potential Glycosylation Sites: N281, N720	MOTIFS
			Tyrosine specific protein phosphatases active site: V409-I421	MOTIFS
51	7525097CD1	592	Protein kinase domain: V116-S394 Src homology 3 domains: G312-D370 Tyrosine kinase, catalytic domain: V116-I540 Receptor tyrosine kinase class V IPB001426: K141-M194, L226-A247, T248-R274, P282-S314 Receptor tyrosine kinase class III IPB001824: E104-G156, L221-K260, G280-V322 Protein kinases signatures and profile: L221-A273	HMMER_PFAM HMMER_SMART HMMER_SMART BLIMPS_BLOCKS BLIMPS_BLOCKS PROFILESCAN

Table 3

SEQ ID	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Tyrosine kinase catalytic domain signature PR00109: M196-T209, Y235-L253, I285-L295, S304-G326	BLIMPS_PRINTS
			KINASE TRANSFERASE ATP-BINDING TYROSINE PROTEIN NON-RECEPTOR ACK PD015776: M1-E113	BLAST_PRODOM
			KINASE ATP-BINDING TRANSFERASE TYROSINE PROTEIN NON-RECEPTOR P21 CDC42HS TYROSINE ACTIVATED PD020576: P499-L589	BLAST_PRODOM
			TYROSINE KINASE PD090676: A311-V455	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM00004 Q00944 423-668: G119-G310, E481-P491 DM00004 A57434 426-671: L121-E340 DM00004 P53356 486-732: G119-G310 DM00004 S33596 183-434: E120-G310	BLAST_DOMO
			Potential Phosphorylation Sites: S60, S96, S139, S149, S255, S294, S302, S445, S513, S563, S566, T33, T258, T337, T355, T367, T532, T543	MOTIFS
			Potential Glycosylation Sites: N31	MOTIFS
			Protein kinases ATP-binding region signature: L122-K148	MOTIFS
			Tyrosine protein kinases specific active-site signature: I241-L253	MOTIFS
52	7525117CD1	118	signal cleavage: M1-R30 KINASE ATP-BINDING TRANSFERASE SERINE/THREONINE PROTEIN TYROSINE PROTEIN AURORA-RELATED STK-1 AIM-1 AIK2 PD012288: M1-D76	SPSCAN BLAST_PRODOM
			Potential Phosphorylation Sites: T73	MOTIFS

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs		Analytical Methods and Databases
			Potential Glycosylation Sites: N60	MOTIFS	
			Protein kinases ATP-binding region signature: L83-K106	MOTIFS	
53	7516593CD1	564	signal_cleavage: M1-V22	SPSCAN	
			POLO box duplicated region: Y417-K480	HMMER_PFAM	
			Protein kinase domain: Y53-F305	HMMER_PFAM	
			Serine/Threonine protein kinases, catalytic domain: Y53-F305	HMMER_SMART	
			Protein kinases signatures and profile: E152-G205	PROFILESCAN	
			KINASE SERINE/THREONINE PROTEIN ATP-BINDING TRANSFERASE NUCLEAR PLK-1 2.7.1. STPK13	BLAST_PRODOM	
			PD026794: M1-R52	BLAST_PRODOM	
			ATP-BINDING TRANSFERASE SER/THR KINASES SERINE/THREONINE PROTEIN SIMILAR	BLAST_PRODOM	
			PD129110: P334-R516	BLAST_PRODOM	
			KINASE PROTEIN SERINE/THREONINE PROTEIN PLK PLK1 SERINE THREONINE STPK13 TRANSFERASE ATP-BINDING	BLAST_PRODOM	
			PD150258: Y481-W514	BLAST_PRODOM	
			PROTEIN KINASE DOMAIN	BLAST_PRODOM	
			DM00004 A54596 55-295: R55-I296		
			DM00004 P52304 27-267: R55-I296		
			DM00004 P53350 55-295: R55-I296		
			POLO-HOMOLOGY DOMAIN	BLAST_PRODOM	
			DM01705 P53350 374-591: L374-Q536	BLAST_PRODOM	
			Potential Phosphorylation Sites: S49, S71, S137, S224, S330, S335, S418, S439, S538, T149, T199, T210, T291, T295, T477, T498	MOTIFS	
			Potential Glycosylation Sites: N385, N437, N527	MOTIFS	

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			MOTIFS	
			Protein kinases ATP-binding region signature: L59-K82	
54	7516603CD1	244	Serine/Threonine protein kinases active-site signature: V172-L184 Guanylate kinase: T40-S144 Guanylate kinase homologues: G3-G189 Guanylate kinase IPB000619-V7-L24, V36-V56, I65-K111 KINASE GUANYLATE TRANSFERASE ATP-BINDING GMP DOMAIN SH3 HOMOLOG MEMBRANE DISCS PD001338: T39-S144	HMMER_PFAM HMMER_SMART BLIMPS_BLOCKS BLAST_PRODOM
			KINASE TRANSFERASE GUANYLATE GMP ATP-BINDING PROTEOME COMPLETE 3D-STRUCTURE ACETYLATION PD003452: L145-E187	BLAST_PRODOM
			GUANYLATE KINASE DM00755 P15454 1-185: R5-E183 DM00755 P21074 1-192: M1-E187 DM00755 P46195 1-194: S2-E187 DM00755 S32545 1-196: P4-E187	BLAST_DOMO
			Potential Phosphorylation Sites: S158, S210, S217, T39, T83, T139 Potential Glycosylation Sites: N171	MOTIFS
			ATP/GTP-binding site motif A (P-loop): G11-S18 Guanylate kinase signature: T39-V56	MOTIFS
55	7525215CD1	698	CNH domain: A486-R696 Protein kinase domain: F15-T272	HMMER_PFAM HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			Domain found in NIK1-like kinases, mouse citron and yeast ROM1. ROM2: A486-R696 Serine/Threonine protein kinases, catalytic domain: F15-T272	HMMER SMART
			Protein kinases signatures and profile: Q111-S163	HMMER SMART
			KINASE SERINE/THREONINE PROTEIN BINDING PHORBOL ESTER ATP-BINDING TRANSFERASE GDP-GTP EXCHANGE RHO1 CDC42-BINDING PD014445: W490-K637	PROFILESCAN
			GC-KINASE RAB8 INTERACTING PROTEIN PD155713: T272-T489	BLAST PRODOM
			PROTEIN KINASE DOMAIN DM000041A53714I17-262: L17-A263 DM000041I49376I270-509: L18-A263 DM000041P08458I20-262: V21-A263 DM000041P35465I271-510: R14-A263	BLAST DOMO
			Potential Phosphorylation Sites: S7, S96, S169, S297, S370, S384, S393, S526, S543, S588, T25, T202, T262, T325, T344, T381, T400, T446, T464, T573, T577	MOTIFS
			Protein Kinases ATP-binding region signature: V21-K44	MOTIFS
56	7525356CD1	486	Protein kinase domain: Y18-F270 Serine/Threonine protein kinases, catalytic domain: Y18-F270	HMMER PFAM
			Protein kinases signatures and profile: Q83-G169	HMMER SMART
			KINASE TRANSFERASE ATP-BINDING SERINE/THREONINE-PROTEIN TYROSINE PROTEIN RECEPTOR 2.7.1. PHOSPHORYLATION PRECURSOR PD000001: I19-V98, M93-S163, M153-F216, L246-P280	PROFILESCAN
			KINASE ATP-BINDING TRANSFERASE SERINE/THREONINE PROTEIN CHAIN TYROSINE PROTEIN ALPHA-2',5'-AMP-ACTIVATED AMPK ACID PD006360: E300-R395	BLAST PRODOM

Table 3

SEQ ID NO.	Incyte Polypeptide ID	Amino Acid Residues	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
			5'-AMP ACTIVATED PROTEIN KINASE, CATALYTIC ALPHA1 CHAIN EC 2.7.1. AMPK TRANSFERASE SERINE/THREONINE PROTEIN KINASE FATTY ACID BIOSYNTHESIS PHOSPHORYLATION ATP-BINDING MULTIGENE FAMILY PD122363: Q419-Q486	BLAST_PRODOM
			PROTEIN KINASE DOMAIN DM000041A53621 18-258: L20-I261 DM000041P54645 17-258: I19-I261 DM000041P54646 18-258: L20-I261 DM000041S51025 18-258: L20-I261	BLAST_DOMO
56, continued			Potential Phosphorylation Sites: S99, S118, S127, S167, S188, S403, S433, S435, S458, T87, T260, T288, T309, T315, T453, T463, Y234	MOTIFS
			Protein kinases ATP-binding region signature: L24-K47	MOTIFS
			Serine/Threonine protein kinases active-site signature: V137-L149	MOTIFS

Table 4

Polymerotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
577521809CB1 1395	1-561, 1-717, 1-903, 2-1152, 463-1395
587520259CB1 1008	1-557, 2-557, 230-1008
597521738CB1 654	1-654, 16-653
607522266CB1 1024	1-872, 167-1024, 324-1024, 334-1024, 335-1024, 338-1023, 338-1024, 378-1024
617523011CB1 952	1-952, 2-723
627523290CB1 1200	1-740, 2-1199, 574-1200
637523379CB1 1162	1-130, 1-765, 2-1161, 350-1162
647523387CB1 733	1-733, 2-732, 86-733, 153-576, 258-733
657521804CB1 1336	1-486, 1-856, 2-721, 547-1336
667521841CB1 978	1-907, 2-668, 201-978
677521886CB1 840	1-840, 2-584
687521897CB1 744	1-527, 1-621, 2-667, 3-742, 3-744
697521995CB1 427	1-427, 2-426, 251-427
707522018CB1 1341	1-870, 641-1340, 641-1341

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
717523799CB1 6470	1-407, 1-420, 1-537, 1-568, 1-605, 1-704, 1-6194, 10-248, 326-706, 385-719, 388-719, 467-716, 467-732, 516-717, 516-723, 524-919, 587-1077, 644-1258, 691-1242, 716-1264, 805-1054, 805-1344, 805-1345, 805-1444, 805-1579, 869-1495, 958-1613, 993-1519, 1268-1816, 1305-2013, 1339-2004, 1375-2177, 1441-1930, 1441-1980, 1441-1991, 1605-2101, 1706-2058, 1768-2177, 1806-2177, 1864-2231, 1891-2478, 2015-2172, 2015-2231, 2015-2387, 2015-2571, 2079-2518, 2237-2513, 2237-2516, 2237-2522, 2237-2524, 2237-2529, 2237-2547, 2237-2551, 2237-2556, 2237-2559, 2237-2560, 2237-2563, 2237-2567, 2237-2572, 2237-2575, 2237-2597, 2237-2604, 2237-2614, 2237-2617, 2237-2624, 2237-2648, 2237-2668, 2237-2672, 2237-2678, 2237-2684, 2237-2689, 2237-2691, 2237-2693, 2237-2700, 2237-2701, 2237-2717, 2237-2729, 2237-2745, 2237-2746, 2237-2753, 2237-2759, 2237-2771, 2237-2795, 2237-2797, 2237-2803, 2237-2828, 2237-2834, 2237-2838, 2237-2841, 2237-2887, 2237-2895, 2237-2981, 2239-2720, 2239-2745, 2240-2510, 2240-2520, 2240-2524, 2241-2688, 2241-2833, 2276-2647, 2280-2863, 2289-2809, 2350-2844, 2362-2440, 2364-2440, 2431-2763, 2471-3189, 2500-3204, 2513-3099, 2550-3188, 2564-3281, 2577-3181, 2599-2820, 2604-2851, 2604-2886, 2604-2917, 2607-3271, 2612-2865, 2613-3232, 2613-3368, 2635-2886, 2644-3093, 2662-3238, 2666-3571, 2701-3162, 2723-3264, 2728-3227, 2749-3518, 2759-3251, 2772-3374, 2775-3300, 2792-2927, 2798-3612, 2809-2855, 2809-2858, 2809-2923, 2809-2927, 2809-2930, 2810-2887, 2810-2891, 2810-2894, 2811-2891, 2811-2927, 2811-3227, 2818-2855, 2818-2887, 2818-2891, 2818-2923, 2827-2855, 2829-2891, 2831-2927, 2838-2888, 2845-2916, 2845-2923, 2845-2931, 2845-2932, 2845-2960, 2845-2963, 2846-2924, 2847-2963, 2854-2955, 2856-2963, 2865-2927, 2865-2963, 2867-3472, 2882-2928, 2882-2952, 2882-2958, 2882-2959, 2882-2960, 2882-2961, 2882-2963, 2882-3613, 2883-2960, 2887-2927, 2889-3227, 2890-2963, 2892-2963, 2899-3299, 2901-2927, 2902-2963, 2903-2963, 2917-2959, 2919-2963, 2923-3512,
	2928-2963, 2931-3606, 2934-3457, 2934-3718, 2937-2963, 2939-3605, 2939-3521, 2939-3645, 2948-3628, 2957-3463, 2966-3752, 2971-3540, 2990-3550, 2990-3796, 2998-3699, 3000-3598, 3010-3612, 3015-3648, 3023-3708, 3026-3666, 3030-3516, 3041-3553, 3070-3652, 3083-3675, 3124-3796, 3140-3787, 3149-3693, 3165-3820, 3173-3986, 3187-3727, 3194-3727, 3204-3803, 3204-3861, 3204-3887, 3204-3896, 3210-3860, 3212-3889, 3213-3984, 3221-3693, 3228-3889, 3228-3903, 3256-3794, 3259-3708, 3262-3990, 3269-3973, 3277-3811, 3278-3875, 3278-3987, 3280-3878, 3297-3995, 3299-4075, 3301-4019, 3301-4071, 3303-3900, 3306-3979, 3310-4075, 3316-3855, 3320-3911, 3334-3911, 3341-3900, 3343-3911, 3350-3995, 3355-4103, 3359-4075, 3367-3896, 3371-3995, 3372-4075, 3381-4094, 3384-4075, 3391-3946, 3393-3990, 3393-3995, 3395-3996, 3398-3995, 3404-3973, 3407-3995, 3419-3974, 3425-3958, 3425-3973, 3441-3948, 3449-3990, 3449-4075, 3462-3995, 3467-4075, 3489-3990, 3494-3995, 3497-3995, 3504-3995,

Table 4

Polynucleotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
71, continued	3505-3995, 3516-3995, 3517-3797, 3523-3995, 3530-3995, 3534-3952, 3536-3952, 3538-3995, 3538-4075, 3539-4075, 3540-3942, 3540-3952, 3541-3984, 3541-4075, 3542-3952, 3543-3995, 3545-3989, 3549-3995, 3557-3995, 3565-3995, 3566-3993, 3570-4075, 3585-3995, 3587-4291, 3593-3990, 3598-3995, 3600-3993, 3617-4075, 3634-3995, 3644-3995, 3645-4075, 3648-4223, 3650-4075, 3653-3995, 3657-3979, 3667-3995, 3677-3995, 3681-3942, 3712-3995, 3730-4075, 3732-3995, 3736-3995, 3738-3995, 3739-3995, 3740-4075, 3747-3995, 3750-3991, 3768-4411, 3770-3995, 3778-3995, 3783-3995, 3791-3990, 3792-4401, 3796-3995, 3804-3995, 3804-4122, 3808-4075, 3808-4117, 3809-3995, 3839-3995, 3840-4411, 3853-3995, 3863-3995, 3865-3990, 3865-3995, 3881-3995, 3892-4012, 3895-3995, 3902-4008, 3908-3995, 3913-3995, 3919-3995, 3929-3995, 3931-4494, 4011-4119, 4011-4205, 4232-4593, 4279-4509, 4279-4635, 4279-4677, 4279-4755, 4279-4759, 4279-4782, 4279-4788, 4279-4797, 4279-4901, 4279-4929, 4279-4937, 4279-4939, 4279-4949, 4279-4957, 4279-5000, 4279-5004, 4279-5023, 4279-5046, 4279-5075, 4282-5067, 4282-5073, 4282-5077, 4283-5106, 4286-4676, 4305-5037, 4312-4930, 4337-4525, 4337-4550, 4337-4554, 4343-4554, 4345-4554, 4366-5090, 4412-5086, 4420-4643, 4491-5173, 4549-5345, 4583-5376, 4594-5368, 4598-5463, 4625-5214, 4640-5063, 4678-5451, 4693-5321, 4721-5365, 4750-5201, 4775-5500, 4793-5637, 4815-5269, 4820-5498, 4824-5674, 4829-5535, 4847-5535, 4858-5126, 4864-5571, 4876-5754, 4877-5615, 4879-5103, 4883-5275, 4894-5532, 4894-5574, 4894-5606, 4906-5545, 4910-5718, 4921-5415, 4926-5371, 4937-5638, 4943-5631, 4949-5576, 4953-5591, 4953-5772, 4953-5782, 4959-5699, 4981-5577, 4981-5745, 4988-5646, 4990-5820, 4991-5650, 4993-5656, 4996-5718, 5001-5742, 5005-5701, 5035-5822, 5036-5714, 5037-5754, 5075-5748, 5088-5731, 5104-5645, 5117-5743, 5120-5695, 5127-5734, 5131-5749, 5191-5903, 5201-5866, 5203-5809, 5205-5689, 5219-5650, 5232-5823, 5256-5864, 5262-5903, 5280-5689, 5281-5744, 5293-5903, 5301-5995, 5309-5859, 5311-5903, 5349-5903, 5352-5903, 5353-5860, 5361-5903, 5365-5903, 5367-5903, 5378-5903, 5381-5896, 5381-5903, 5397-5903, 5399-5650, 5403-5903, 5419-5903, 5440-5903, 5451-5837, 5470-5903, 5473-5903, 5480-6014, 5481-5903, 5489-5903, 5490-5903, 5509-5882, 5509-5903, 5517-5903, 5528-5786, 5530-5903, 5531-5789, 5533-5903, 5538-5762, 5545-5803, 5545-5903, 5551-5903, 5576-5903, 5581-5903, 5593-5903, 5595-5903, 5608-5903, 5610-5903, 5615-5898, 5617-5883, 5617-5903, 5620-5903, 5627-5677, 5634-6194, 5638-5903, 5641-5863, 5641-5903, 5644-5922, 5647-5903, 5659-5903, 5682-5794, 5689-5903, 5757-5903, 5780-5903, 5827-6468, 5852-5903, 5860-5903, 5895-6015, 5901-6015, 5901-6238, 5901-6239, 5901-6348, 5901-6351, 5902-6351, 5903-5955, 5903-5998, 5903-6116, 5903-6129, 5903-6136, 5903-6157, 5903-6168, 5903-6238, 5903-6282, 5903-6336, 5903-6355, 5903-6411, 5903-6428, 5903-6463, 5903-6465, 5903-6468, 5903-6470, 5904-5955, 5924-6462, 5943-6468, 5945-6468, 5958-6422, 5960-6422, 5984-6251, 6013-6351, 6054-6468, 6064-6468, 6074-6468, 6086-6468, 6092-6468, 6094-6292, 6127-6468

Table 4

Polymer ID SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
72/7521743CB1 1220	1-676, 402-1220, 414-1218, 426-1220, 930-1220
73/7522317CB1 1084	1-771, 2-759, 169-1084
74/7522400CB1 834	1-113, 1-466, 1-798, 1-834, 2-796
75/7523524CB1 1406	1-751, 2-735, 640-1406
76/7523542CB1 1640	1-922, 2-1639, 258-1099, 315-1130, 612-1561, 645-1383, 944-1640, 1057-1640
77/7523546CB1 1810	1-717, 2-1805, 289-1240, 308-1252, 314-1259, 327-1261, 547-1221, 562-1329, 568-1327, 568-1377, 922-1806, 1040-1810
78/7523552CB1 1484	1-574, 1-782, 1-807, 2-1483, 477-1360, 477-1395, 679-1484
79/7523564CB1 1675	1-621, 1-776, 1-777, 1-945, 2-1674, 56-925, 56-963, 696-1626, 699-1552, 699-1675, 704-1476, 845-1675, 849-1675, 1211-1675
80/7523572CB1 1489	1-621, 1-776, 1-777, 1-918, 1-980, 2-1475, 84-984, 692-1489, 697-1489, 713-1489
81/7523586CB1 1775	1-823, 2-1775, 197-1090, 270-1097, 657-1297, 662-1389, 963-1775
82/7523617CB1 2776	1-818, 2-2775, 622-1377, 634-1501, 1405-2306, 1548-2310, 2144-2776, 2147-2776, 2195-2776, 2212-2776, 2226-2776
83/7523625CB1 1683	1-803, 2-1682, 428-1310, 662-1465, 667-1630, 1223-1683
84/7523650CB1 2020	1-701, 1-702, 1-715, 1-749, 1-837, 1-846, 2-1925, 613-1346, 627-1380, 636-1505, 692-1352, 1156-2020
85/7523665CB1 1369	1-855, 1-1368, 5-295, 170-847, 170-851, 199-874, 545-1369, 790-1369

Table 4

Polymer SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
86/7523672CB1 1759	1-590, 1-611, 1-612, 1-641, 1-643, 1-704, 1-730, 1-741, 1-794, 1-936, 2-1758, 319-1095, 435-1066, 665-1512, 667-1458, 911-1759
87/7523687CB1 2480	1-880, 2-2479, 689-1264, 755-1253, 859-1826, 939-1800, 1645-2480
88/7523689CB1 1828	1-505, 1-751, 1-924, 2-1827, 341-1196, 467-1199, 691-1503, 691-1636, 1024-1828, 1032-1824
89/7523705CB1 1505	1-842, 3-1504, 130-871, 183-877, 669-1505, 689-1505, 735-1505
90/7523706CB1 1334	1-571, 1-796, 1-804, 2-1334, 473-1314, 473-1326, 557-1334, 615-1334, 659-1330
91/7523707CB1 2234	1-876, 2-1984, 297-1288, 461-1308, 687-1605, 708-1576, 1106-1984, 1669-2234
92/7523719CB1 1696	1-630, 1-715, 1-733, 2-1695, 525-1450, 555-1245, 555-1282, 555-1304, 1047-1696, 1121-1696, 1152-1696
93/7523720CB1 2432	1-716, 1-808, 1-911, 2-2432, 590-1358, 590-1545, 879-1824, 926-1862, 976-1826, 1003-1831, 1744-2432, 1746-2432, 1794-2432
94/7523737CB1 2009	1-701, 1-702, 1-715, 1-749, 1-854, 1-921, 1-935, 2-2008, 481-1337, 546-1336, 619-1374, 629-1284, 1262-2009
95/7523742CB1 1711	1-689, 1-762, 1-824, 2-1710, 5-685, 140-1080, 185-1091, 371-1091, 379-1101, 880-1711, 882-1711, 888-1711, 893-1711, 983-1711
96/7523743CB1 1677	1-630, 1-754, 1-912, 2-1676, 463-1347, 464-1463, 824-1677, 897-1677, 1019-1677
97/7523745CB1 1876	1-859, 2-1875, 375-1304, 395-1300, 597-1348, 633-1464, 945-1876, 1012-1876, 1028-1876, 1065-1876
98/7523757CB1 2363	1-729, 2-2362, 659-1345, 675-1469, 844-1719, 932-1728, 1513-2363
99/7523770CB1 2032	1-795, 1-962, 2-2031, 45-548, 45-794, 45-955, 524-1367, 539-1346, 736-1666, 737-1561, 1113-2032, 1146-2032, 1243-2032

Table 4

Polymerotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
100/7523919CB1 2299	1-177, 1-925, 2-2289, 2-2298, 638-1452, 642-1462, 672-1568, 698-1583, 1510-2299, 1677-2298
101/7522140CB1 1792	1-1792, 882-1425, 882-1451, 882-1470, 882-1474, 882-1476, 950-1009, 950-1397, 950-1505, 950-1618, 1559-1711
102/7522525CB1 1365	1-726, 1-1365, 396-1365, 416-1365, 419-1365, 422-1365, 426-1365, 456-1365, 472-1365, 473-1365, 535-1361, 563-1365, 578-1364, 594-1365
103/7523355CB1 2554	1-778, 1-918, 2-2554, 5-777, 695-1456, 702-1644, 968-1877, 1045-1878, 1648-2554, 1682-2554, 1808-2554, 1890-2554
104/7524443CB1 4386	1-804, 1-807, 1-839, 53-4386, 362-1141, 401-1273, 744-1556, 744-1614, 753-1634, 769-1535, 770-1577, 783-1761, 1187-1999, 1194-2164, 1576-2433, 1597-2432, 2257-3084, 2267-3100, 2896-3517, 2747-3547, 2803-3675, 2809-3678, 3321-4038, 3329-4206, 3521-4386, 3526-4385, 3616-4386, 3683-4386, 3693-4386, 3708-4386, 4097-4386, 4107-4386
105/7524498CB1 3924	1-765, 1-784, 1-800, 52-3924, 362-1142, 400-1238, 687-1353, 688-1540, 746-1619, 783-1774, 1194-1941, 1195-2089, 1581-2236, 1587-2524, 2009-2772, 2037-2903, 2374-3258, 2427-3224, 2427-3291, 2582-3244, 2859-3623, 2879-3744, 3038-3924, 3067-3924, 3140-3917, 3154-3922, 3237-3924, 3249-3924, 3259-3924, 3654-3924, 3679-3924
106/7524957CB1 3770	1-577, 43-83, 43-3770, 59-83, 108-852, 108-885, 108-886, 108-897, 172-472, 172-855, 175-821, 184-391, 361-1294, 366-1009, 506-1371, 509-1044, 675-1211, 682-1404, 682-1450, 688-1590, 710-1405, 714-972, 725-1407, 725-1472, 732-1409, 739-1668, 740-1406, 740-1618, 853-1282, 853-1451, 864-1407, 925-1413, 1024-1621, 1035-1631, 1050-1661
107/7525097CB1 1836	1-616, 1-626, 1-647, 1-662, 1-680, 1-730, 1-776, 346-1213, 480-1215, 594-1305, 605-1487, 982-1836, 1075-1836, 1120-1836, 1141-1836
108/7525117CB1 960	1-802, 2-959, 173-960, 340-960

Table 4

Polymerotide SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
109/7516593CB1 2275	1-209, 1-227, 1-271, 1-282, 1-309, 1-439, 1-2275, 3-414, 5-763, 9-217, 14-735, 15-249, 15-368, 15-410, 15-443, 15-468, 15-478, 15-495, 15-534, 15-546, 15-550, 15-688, 15-768, 15-779, 15-858, 18-271, 20-252, 25-332, 27-317, 28-591, 33-608, 36-280, 43-332, 44-271, 49-310, 50-238, 56-271, 57-271, 63-281, 63-296, 65-243, 65-700, 65-760, 66-300, 66-302, 68-930, 71-324, 72-555, 82-316, 85-271, 171-464, 175-548, 211-915, 217-715, 224-674, 264-797, 277-517, 278-758, 298-560, 301-700, 301-937, 355-538, 358-863, 363-872, 369-886, 374-557, 384-741, 410-532, 481-928, 562-1100, 563-1111, 565-1159, 603-1134, 621-897, 625-840, 670-1157, 671-1265, 676-937, 703-1448, 738-1341, 747-914, 771-1029, 791-1344, 792-1466, 799-1095, 819-1102, 819-1123, 827-1408, 858-1489, 866-1489, 892-1414, 899-1655, 899-1658, 899-1664, 899-1667, 906-1441, 921-1504, 948-1241, 948-1272, 948-1291, 949-1118, 964-1204, 971-1289, 979-1474, 998-1273, 998-1296, 1002-1155, 1003-1305, 1003-1505, 1003-1596, 1003-1666, 1009-1280, 1009-1407,
1009/1409, 1014-1522, 1043-1542, 1047-1323, 1053-1667, 1060-1302, 1065-1242, 1076-1627, 1090-1339, 1095-1465, 1096-1339, 1096-1508, 1096-1562, 1096-1622, 1114-1613, 1117-1375, 1132-1388, 1151-1602, 1167-1382, 1167-1401, 1172-1278, 1172-1329, 1172-1449, 1172-1618, 1172-1631, 1172-1726, 1180-1420, 1211-2040, 1223-1622, 1278-1555, 1303-1558, 1355-1667, 1361-1681, 1377-1642, 1477-1639, 1542-1667, 1760-1796, 1763-2275, 1790-2045, 1790-2190, 1791-2269, 1791-2275, 1795-2270, 1804-2275, 1806-2269, 1808-2207, 1823-2268, 1829-2275, 1831-2275, 1832-2275, 1836-1960, 1837-2269, 1850-2270, 1852-2250, 1865-2129, 1865-2134, 1867-2267, 1876-2198, 1876-2274, 1876-2275, 1877-2273, 1878-2139, 1881-2270, 1909-2223, 1909-2269, 1912-2270, 1916-2275, 1924-2267, 1933-2270, 1938-2270, 1939-2270, 1945-2270, 1947-2234, 1950-2275, 1962-2270, 1964-2270, 1978-2275, 1979-2266, 1989-2269, 1991-2253, 1993-2106, 1998-2269, 2006-2269, 2010-2269, 2018-2270, 2028-2271, 2030-2270, 2060-2269, 2070-2270, 2080-2275, 2087-2275, 2147-2270, 2148-2275, 2161-2270, 2169-2270, 2173-2271, 2178-2269	
110/7516603CB1 1168	1-277, 1-1128, 147-308, 192-591, 193-433, 193-494, 193-520, 194-354, 194-410, 194-440, 194-460, 194-462, 194-472, 194-542, 194-545, 194-555, 194-564, 194-612, 194-613, 194-623, 194-631, 194-633, 194-637, 194-643, 194-646, 194-649, 194-652, 194-658, 194-671, 194-672, 194-756, 195-619, 196-403, 196-636, 197-407, 197-646, 197-655, 199-468, 199-486, 199-726, 200-379, 201-445, 201-446, 201-447, 205-475, 205-556, 207-457, 210-575, 213-552, 214-411, 216-630, 220-500, 225-349, 225-588, 230-578, 231-585, 232-500, 236-496, 240-444, 242-431, 242-454, 242-455, 242-647, 243-489, 243-677, 246-466, 249-508, 250-507, 252-530, 262-454, 264-518, 264-522, 265-758, 288-547, 268-510, 272-542, 288-532, 289-541, 289-549, 289-552, 289-583, 290-573, 294-628, 298-506, 300-563, 304-504, 304-539, 304-557, 304-598, 306-540, 311-530, 311-581, 314-556, 315-564, 315-582, 318-544, 323-521, 323-585, 324-542, 324-555, 326-580, 326-581, 327-540, 331-598, 332-610, 333-591, 336-610, 336-628, 337-589, 339-637, 340-572, 344-

Table 4

Polymer SEQ ID NO./ Incyte ID/ Sequence Length	Sequence Fragments
350-587, 352-614, 353-683, 356-549, 358-615, 366-617, 366-651, 367-699, 368-616, 368-624, 369-616, 370-625, 370-641, 372-535, 373-	
758, 375-623, 376-627, 378-583, 378-621, 378-637, 379-603, 379-613, 379-621, 379-629, 381-567, 382-602, 382-620, 382-643,	
383-644, 384-662, 390-613, 392-602, 398-758, 399-616, 399-671, 401-683, 402-541, 402-668, 402-674, 403-687, 407-674, 409-	
601, 409-613, 415-544, 417-681, 418-696, 419-681, 421-651, 422-647, 425-669, 425-694, 426-704, 427-644, 427-646, 427-670, 427-688,	
428-581, 428-699, 434-698, 435-630, 436-674, 437-640, 440-665, 441-692, 441-718, 442-663, 445-635, 446-720, 447-715, 450-585, 452-	
711, 452-724, 455-726, 459-749, 464-668, 464-675, 466-751, 475-693, 480-694, 480-745, 482-737, 489-729, 502-758, 508-758, 517-758,	
519-740, 520-758, 532-733, 535-758, 536-724, 541-758, 553-749, 567-674, 622-1133, 636-758, 651-758, 724-966, 836-875, 836-891, 836-	
921, 848-1119, 848-1120, 850-1123, 851-1119, 853-1119, 865-1109, 874-1115,	
880-1134, 895-1115, 904-1141, 905-1136, 905-1168, 906-1134, 915-1167, 918-1165, 929-1133, 995-1126, 997-1079, 1050-1119	
111/7525215CB1 1-628, 1-731, 1-777, 2-554, 2-2559, 95-2559, 698-1444, 702-1511, 706-1668, 718-1610, 1485-2335, 1745-2560, 1750-2560, 1956-2534,	
2560 1957-2560	
112/7525356CB1 1-122, 1-443, 1-465, 1-545, 1-572, 1-611, 2-513, 11-264, 16-628, 24-273, 25-254, 25-449, 498-780, 539-727, 646-881, 682-904, 691-890,	
1662 691-1290, 691-1322, 691-1324, 1104-1662	

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
71	7523799CB1	BRAIFFE02
91	7523707CB1	BONTNOT01
106	7524957CB1	LIVRTUE01
109	7516593CB1	HNT2RAT01
110	7516603CB1	COLNNOT19
112	7525556CB1	PROSNOT01

Table 6

Library	Vector	Library Description
BONTNOT01	pINCY	Library was constructed using RNA isolated from tibial periosteum removed from a 20-year-old Caucasian male during a hemipelvectomy with amputation above the knee. Pathology for the associated tumor tissue indicated partially necrotic and cystic osteoblastic grade 3 osteosarcoma (post-chemotherapy). Family history included osteogenesis imperfecta, closed fracture, and type II diabetes.
BRAIFEF02	PCMV-ICIS	This full-length enriched library was constructed using RNA isolated from brain tissue removed from a Caucasian male fetus who was stillborn with a hypoplastic left heart at 23 weeks' gestation. Serologies were negative.
COLNNOT19	pINCY	Library was constructed using RNA isolated from the cecal tissue of an 18-year-old Caucasian female. The cecal tissue, along with the appendix and ileum tissue, were removed during bowel anastomosis. Pathology indicated Crohn's disease of the ileum involving 15 cm of the small bowel.
HNT2RAT01	PBLUESCRIPT	Library was constructed at Stratagene (STR937231), using RNA isolated from the hNT2 cell line (derived from a human teratocarcinoma that exhibited properties characteristic of a committed neuronal precursor). Cells were treated with retinoic acid for 24 hours.
LIVRTUE01	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from liver tumor tissue removed from a 72-year-old Caucasian male during partial hepatectomy. Pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father.
PROSNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the prostate tissue of a 78-year-old Caucasian male, who died from leukemia. Patient history included skin cancer, emphysema, and asthma. Previous surgeries included a cholecystectomy.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	<i>ESTs</i> : Probability value=1.0E-8 or less <i>Assembled ESTs</i> : fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : Probability value=1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises at least five functions: fasta, tfasta, fastx, tblastx, and sssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad. Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	<i>ESTs</i> : fasta E value=1.0E-6 <i>Assembled ESTs</i> : fasta Identity=95% or greater and Match length=200 bases or greater; fastx E value=1.0E-8 or less <i>Full Length sequences</i> : fastx score=100 or greater
BLIMPS	A BLOCKs IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. & S. Henikoff (1996) Methods Enzymol. 266:88-105; and Altwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value=1.0E-3 or less
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART, and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1998) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, p. 1-350	<i>PFAM, INCY, SMART, or TIGRFAM hits</i> : Probability value=1.0E-3 or less <i>Signal peptide hits</i> : Score= 0 or greater

Table 7 (cont.)

Program	Description	Reference	Parameter Threshold
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score≥GCG-specified "HIGH" value for that particular Prosite motif. Generally, score=1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phred Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score=120 or greater; Match length=56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score=3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Int'l. Conf. on Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence Press, Menlo Park, CA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
57	7521809	7155864H1	SNP00117534	371	984	T	C	T	noncoding	0.58	0.29	0.51
59	7521738	080882H1	SNP0054419	191	586	A	A	G	N182	n/a	n/a	n/a
59	7521738	1301819F6	SNP0054419	181	587	A	A	G	K182	n/a	n/a	n/a
59	7521738	1315762F6	SNP0054419	110	589	A	A	G	N183	n/a	n/a	n/a
59	7521738	2517370F6	SNP00101383	379	206	G	G	A	R55	0.93	n/a	n/a
59	7521738	2525814F7	SNP00101383	396	205	A	G	A	Q55	0.93	n/a	n/a
59	7521738	3086461H1	SNP00153284	166	441	G	G	A	G134	n/a	n/a	n/a
59	7521738	3424368H1	SNP00101383	204	207	G	G	A	A56	0.93	n/a	n/a
59	7521738	7335330H1	SNP00101383	387	208	A	G	A	D56	0.93	n/a	n/a
61	7523011	2534938H2	SNP00024525	218	215	G	G	C	A51	n/d	n/d	n/d
62	7523290	2313881T6	SNP00069599	407	762	C	C	A	noncoding	0.75	0.78	0.7
62	7523290	2837962T6	SNP00003592	27	1141	T	T	C	noncoding	n/a	n/a	n/a
62	7523290	405345H1	SNP00003592	190	1123	T	T	C	noncoding	n/a	n/a	n/a
62	7523290	5313586F6	SNP00154287	252	830	G	A	G	noncoding	n/a	n/a	n/a
62	7523290	760934T7	SNP00154287	331	849	A	A	G	noncoding	n/a	n/a	n/a
63	7523379	1404269F6	SNP00027149	113	402	T	C	T	noncoding	n/d	0.99	0.98
63	7523379	2510275F6	SNP00126667	546	1092	A	A	G	noncoding	n/a	n/a	n/a
63	7523379	3092795H1	SNP00013704	78	743	C	C	T	noncoding	n/a	n/a	n/a
64	7523387	1582436H1	SNP00132581	138	14	A	A	G	K5	n/a	n/a	n/a
64	7523387	1816291T6	SNP00016682	324	553	C	C	T	N184	n/d	n/a	n/a
65	7521804	244804H1	SNP0000977	132	1296	G	A	G	noncoding	n/a	n/a	n/a
66	7521841	27538714H1	SNP00104709	24	748	G	G	A	K249	n/d	n/a	n/a
66	7521841	4052503H1	SNP00136019	143	685	C	C	T	G228	n/a	n/a	n/a
66	7521841	6907635H1	SNP00123039	495	594	A	A	G	H198	n/a	n/a	n/a
66	7521841	7021645H1	SNP00123040	159	684	G	T	G	G228	n/a	n/a	n/a
66	7521841	7633346J1	SNP00136018	148	51	A	A	C	H17	n/a	n/a	n/a
69	7521995	2052218T6	SNP00036439	146	427	G	G	A	noncoding	0.97	n/a	n/a
69	7521995	2159517T6	SNP00128159	243	312	T	T	C	noncoding	n/a	n/a	n/a

Table 8

SEQ ID NO.	PD	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
69	7521995	262741H1	SNP00036439	307	426	A	G	A	noncoding	0.97	n/a	n/a
69	7521995	262741R6	SNP00036439	307	425	A	G	A	noncoding	0.97	n/a	n/a
69	7521995	2735091T6	SNP00128159	300	262	T	T	C	noncoding	n/a	n/a	n/a
69	7521995	3539442H1	SNP0008812	162	256	G	A	G	noncoding	n/a	n/a	n/a
71	7523799	1227227R6	SNP00005102	189	6376	G	G	T	noncoding	0.86	0.74	0.99
71	7523799	1227227R6	SNP00026859	134	6321	C	C	T	noncoding	n/a	n/a	n/a
71	7523799	1227227T6	SNP00005102	37	6375	G	G	T	noncoding	0.86	0.74	0.99
71	7523799	1227227T6	SNP00026859	92	6320	C	C	T	noncoding	n/a	n/a	n/a
71	7523799	1417295T6	SNP00005102	13	6405	G	G	T	noncoding	0.86	0.74	0.99
71	7523799	1417295T6	SNP00026859	68	6350	C	C	T	noncoding	n/a	n/a	n/a
71	7523799	1503744F6	SNP00014536	47	5265	C	C	T	P1748	n/a	n/a	n/a
71	7523799	1503744F6	SNP00014537	77	5295	C	C	T	D1758	0.69	0.68	0.77
71	7523799	1503744F6	SNP00014538	228	5446	C	C	T	Q1809	n/a	n/a	n/a
71	7523799	2949625F6	SNP00103259	311	835	A	A	G	N272	n/a	n/a	n/a
71	7523799	2949625T6	SNP00103259	342	881	A	A	G	stop287	n/a	n/a	n/a
71	7523799	2949625T6	SNP00103260	185	1038	C	T	C	F339	0.95	n/a	n/a
71	7523799	3793262H1	SNP00130998	200	5814	C	C	A	R1931	n/a	n/a	n/a
71	7523799	6764248H1	SNP00111622	386	2461	A	A	G	M814	0.47	n/a	n/a
71	7523799	6823369J1	SNP00111623	15	1905	T	T	C	D628	0.57	n/a	n/a
71	7523799	6888512J1	SNP00103260	123	964	T	T	C	F315	0.95	n/a	n/a
71	7523799	7100879H1	SNP00103259	119	834	G	A	G	T271	n/a	n/a	n/a
71	7523799	7100879H1	SNP00103260	275	990	T	T	C	S323	0.95	n/a	n/a
72	7521743	1858957F6	SNP00025472	113	852	C	T	C	noncoding	n/d	n/a	n/a
72	7521743	1913432H1	SNP00025472	69	851	T	T	C	noncoding	n/d	n/a	n/a
73	7522317	1390895F6	SNP00146386	198	124	G	G	A	R37	n/a	n/a	n/a
73	7522317	4347211F6	SNP00146387	129	612	C	C	T	noncoding	n/a	n/a	n/a
73	7522317	4347211H1	SNP00143662	199	683	A	A	C	noncoding	n/a	n/a	n/a
74	7522400	159423R6	SNP00114371	218	85	G	G	A	E28	0.8	0.63	0.88

Table 8

SEQ ID NO.	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
75	7523524	122018F1	SNP00021335	137	1343	C	T	H445	n/d	n/d	n/d	n/d
75	7523524	1309709F6	SNP00139670	18	815	C	C	D269	n/a	n/a	n/a	n/a
75	7523524	1434184H1	SNP00148163	12	585	T	T	F193	n/a	n/a	n/a	n/a
75	7523524	2205890T6	SNP00021335	48	1357	C	T	A450	n/d	n/d	n/d	n/d
75	7523524	7632184J1	SNP00148163	175	569	C	T	V187	n/a	n/a	n/a	n/a
76	7523542	4539781F6	SNP00115147	445	818	C	C	Q273	n/d	n/a	n/a	n/a
78	7523552	1746045H1	SNP00065854	169	717	G	G	A noncoding	n/a	n/a	n/a	n/a
78	7523552	2527637H1	SNP00105205	288	795	C	C	A noncoding	n/d	n/a	n/a	n/a
78	7523552	3033264H1	SNP00105206	163	900	C	C	A noncoding	n/d	n/a	n/a	n/a
78	7523552	3238712F6	SNP00071050	258	785	A	G	A noncoding	n/a	n/a	n/a	n/a
78	7523552	3785064H1	SNP00105207	112	1374	T	T	C noncoding	n/d	n/d	n/d	n/a
78	7523552	7693593J2	SNP00105207	300	1372	T	T	C noncoding	n/d	n/d	n/d	n/a
79	7523564	069124H1	SNP00028779	21	150	T	T	G V50	n/d	n/a	n/a	n/a
79	7523564	4407977H1	SNP00122800	124	1317	C	C	T noncoding	n/a	n/a	n/a	n/a
80	7523572	069124H1	SNP00028779	21	150	T	T	G V50	n/d	n/a	n/a	n/a
80	7523572	5597287H1	SNP00122800	142	1131	C	C	T A377	n/a	n/a	n/a	n/a
81	7523586	6420674H1	SNP00155160	296	850	G	G	A noncoding	n/a	n/a	n/a	n/a
82	7523617	4173045F6	SNP00072832	40	2216	T	C	F731	n/d	0.99	n/d	n/d
83	7523625	055029H1	SNP0035691	30	876	T	C	S288	n/a	n/a	n/a	n/a
84	7523650	1831864H1	SNP00015291	94	1658	T	C	V542	0.98	n/a	n/a	n/a
84	7523650	3660453H1	SNP00015290	266	750	G	A	T239	0.7	n/a	n/a	n/a
84	7523650	7741120H1	SNP00106556	42	396	C	C	A121	n/a	n/a	n/a	n/a
84	7523650	7741120H1	SNP00142890	41	395	C	C	A121	n/a	n/a	n/a	n/a
86	7523672	2836716H1	SNP00042776	130	783	T	C	T Y261	n/d	n/a	n/a	n/a
87	7523687	1746456H1	SNP00149665	25	2181	C	C	noncoding	n/a	n/a	n/a	n/a
87	7523687	1746640T6	SNP00149665	274	2188	C	C	noncoding	n/a	n/a	n/a	n/a
88	7523689	1635349F6	SNP00016799	331	1823	C	T	noncoding	n/a	n/a	n/a	n/a
90	7523706	1349640H1	SNP00041139	140	428	T	C	T D141	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PD	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
90	7523706	3128484F6	SNP00115525	247	27	T	T	F8	n/d	n/d	n/d	n/d
90	7523706	3128484F6	SNP00132584	267	47	T	T	N14	n/a	n/a	n/a	n/a
90	7523706	772563311	SNP00115525	508	16	T	T	M4	n/d	n/d	n/d	n/d
90	7523706	772563311	SNP00132584	528	36	C	T	L11	n/a	n/a	n/a	n/a
91	7523707	1714426H1	SNP00074229	74	227	G	G	R68	0.18	0.04	0.03	0.05
91	7523707	6865402H1	SNP00074230	356	1012	T	C	L330	0.77	0.63	0.6	0.65
92	7523719	2451420F6	SNP00115146	276	270	G	G	G90	n/d	n/a	n/d	n/d
92	7523719	4539781F6	SNP00115147	445	1010	C	C	T	Q337	n/d	n/a	n/a
92	7523719	6082148F8	SNP00133611	19	73	G	G	A	P24	n/a	n/a	n/a
92	7523719	6082148F8	SNP00133612	39	93	A	A	G	E31	n/a	n/a	n/a
93	7523720	045705H1	SNP00032102	17	1839	C	C	A	G586	n/a	n/a	n/a
93	7523720	3465726H1	SNP00114230	211	1642	T	C	T	F521	n/d	n/a	n/a
93	7523720	3718481H1	SNP00149710	15	1761	A	A	C	S560	n/a	n/a	n/a
93	7523720	6266287H1	SNP00032101	187	754	C	C	G	P225	n/a	n/a	n/a
93	7523720	7251618F8	SNP00114229	36	549	A	A	G	L156	0.6	0.76	0.55
94	7523737	3666453H1	SNP00015290	266	750	G	G	A	T239	0.7	n/a	n/a
94	7523737	7035770H1	SNP00015291	96	1838	C	C	T	A602	0.98	n/a	n/a
94	7523737	7741120H1	SNP00106336	42	396	C	C	T	A121	n/a	n/a	n/a
94	7523737	7741120H1	SNP00142890	41	395	C	C	G	A121	n/a	n/a	n/a
96	7523743	2451420F6	SNP00115146	276	270	G	G	A	G90	n/d	n/d	n/d
96	7523743	4539781F6	SNP00115147	445	1010	C	C	T	Q337	n/d	n/a	n/a
96	7523743	6082148F8	SNP00133611	19	73	G	G	A	P24	n/a	n/a	n/a
96	7523743	6082148F8	SNP00133612	39	93	A	A	G	E31	n/a	n/a	n/a
99	7523770	1635349F6	SNP00016799	331	2031	C	C	T	noncoding	n/a	n/a	n/a
100	7523919	045705H1	SNP00032102	17	1572	C	C	A	noncoding	n/a	n/a	n/a
100	7523919	3465726H1	SNP00114230	211	1375	T	C	T	noncoding	n/d	n/a	n/a
100	7523919	3718481H1	SNP00149710	15	1494	A	A	C	noncoding	n/a	n/a	n/a
100	7523919	62660287H1	SNP00032101	187	487	C	C	G	P136	n/a	n/a	n/a

Table 8

SEQ ID NO.	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
100	7523919	7251618F8	SNP00114229	36	282	A	A	G	L67	0.6	0.76	0.55
101	7522140	4541894H1	SNP00056374	64	702	T	C	noncoding	n/d	n/a	n/a	0.49
102	7522525	1441018F7	SNP00016048	454	1116	T	C	C360	n/a	n/a	n/a	n/a
102	7522525	1449084R1	SNP00016049	80	1245	T	T	R403	n/a	n/a	n/a	n/a
102	7522525	1706132T6	SNP00016049	307	1252	T	C	S406	n/a	n/a	n/a	n/a
102	7522525	2295842T6	SNP00016048	144	1117	C	C	R361	n/a	n/a	n/a	n/a
102	7522525	2372209T6	SNP00016049	284	1270	T	T	F412	n/a	n/a	n/a	n/a
104	7524443	1741505T6	SNP00054334	116	3258	G	G	A	R1069	n/d	n/d	n/d
104	7524443	1741505T6	SNP00124224	52	3322	T	T	C	S1090	n/d	n/d	n/d
104	7524443	1852144T6	SNP00029583	89	4303	C	C	T	noncoding	n/d	n/a	n/a
104	7524443	2086173H1	SNP00029583	150	4286	C	C	T	noncoding	n/d	n/a	n/a
104	7524443	2103173R6	SNP00074035	137	331	A	A	G	G93	n/d	n/d	n/d
104	7524443	2172376F6	SNP00074035	62	327	A	A	G	K92	n/d	n/d	n/d
104	7524443	2230058H1	SNP00029582	57	3494	C	C	T	H1148	n/d	n/a	n/a
104	7524443	2502887T6	SNP00029583	66	4326	C	C	T	noncoding	n/d	n/a	n/a
104	7524443	2606210F6	SNP00029582	412	3493	C	C	T	L1147	n/d	n/a	n/a
104	7524443	2606210F6	SNP00124225	348	3429	A	A	G	K1126	n/a	n/a	n/a
104	7524443	2606210H1	SNP00054332	20	3102	G	G	A	G1017	n/a	n/a	n/a
104	7524443	2606210H1	SNP00054333	135	3217	G	G	A	R1055	n/d	n/a	n/a
104	7524443	2827761H1	SNP00124223	98	644	G	G	A	E198	n/d	n/d	n/d
104	7524443	3136587H1	SNP00124225	104	3430	A	A	G	T1126	n/a	n/a	n/a
104	7524443	4249977R6	SNP00098419	351	2350	C	C	A	S766	0.86	n/d	n/d
104	7524443	5971646H1	SNP00074037	419	2272	C	G	C	S740	n/a	n/a	n/a
104	7524443	7692816I1	SNP00074038	222	2274	T	T	C	L741	n/a	n/a	n/a
105	7524498	1741505T6	SNP00054334	116	2796	G	G	A	R915	n/d	n/d	n/d
105	7524498	1741505T6	SNP00124224	52	2860	T	T	C	S936	n/d	n/d	n/d
105	7524498	1852144T6	SNP00029583	89	3841	C	C	T	noncoding	n/d	n/a	n/a
105	7524498	2086173H1	SNP00029583	150	3824	C	C	T	noncoding	n/d	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
105	7524498	2103173R6	SNP00074035	137	331	A	A	G	R93	n/d	n/d	n/d	n/d
105	7524498	2172576F6	SNP00074035	62	327	A	A	G	K92	n/d	n/d	n/d	n/d
105	7524498	2230058H1	SNP00029582	57	3032	C	C	T	H94	n/d	n/a	n/a	n/a
105	7524498	2502887T6	SNP00029583	66	3864	C	C	T	noncoding	n/d	n/a	n/a	n/a
105	7524498	2606210F6	SNP00029582	412	3031	C	C	T	L993	n/d	n/a	n/a	n/a
105	7524498	2606210F6	SNP00124225	348	2967	A	A	G	K972	n/a	n/a	n/a	n/a
105	7524498	2606210H1	SNP00054332	20	2640	G	G	A	G863	n/a	n/a	n/a	n/a
105	7524498	2606210H1	SNP00054333	135	2755	G	G	A	R901	n/d	n/a	n/a	n/a
105	7524498	2827761H1	SNP00124223	98	644	G	G	A	E198	n/d	n/d	n/d	n/d
105	7524498	3136587H1	SNP00124225	104	2968	A	A	G	T972	n/a	n/a	n/a	n/a
105	7524498	5971646H1	SNP00074037	419	2011	C	C	C	S653	n/a	n/a	n/a	n/a
105	7524498	6286237H2	SNP00074036	122	1396	G	G	A	E448	n/d	n/d	n/d	n/d
105	7524498	7367225H1	SNP00098419	491	2089	A	C	A	S679	0.86	n/d	n/d	n/d
106	7524957	2890336F6	SNP00100443	263	2727	T	T	C	S856	n/d	n/d	n/d	n/d
106	7524957	4140043H1	SNP00100444	75	2911	G	G	A	V918	n/d	n/a	n/a	n/a
106	7524957	4761874F6	SNP00100444	54	2910	G	G	A	M917	n/d	n/a	n/a	n/a
106	7524957	4761874F6	SNP00100445	318	3174	T	T	C	P1005	n/d	n/d	n/d	n/d
106	7524957	6287317H2	SNP00100445	115	3175	T	T	C	S1006	n/d	n/d	n/d	n/d
106	7524957	7098159H1	SNP00100441	402	1009	A	A	G	S284	0.88	n/a	n/a	n/a
106	7524957	7098159H1	SNP00100442	350	1061	T	T	G	V301	0.57	0.71	0.4	0.39
108	7525117	161207T6	SNP00122098	168	866	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7525117	1679482F6	SNP00122098	272	810	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7525117	1679482T6	SNP00122098	193	845	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7525117	1679812T6	SNP00122098	226	819	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7525117	2758503H1	SNP00122098	183	809	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7525117	3958001T9	SNP00153975	227	674	C	C	T	noncoding	n/a	n/a	n/a	n/a
108	7525117	3958001T9	SNP00153976	84	817	C	C	T	noncoding	n/a	n/a	n/a	n/a
109	7516593	1272419H1	SNP00016248	173	187	C	C	T	V43	n/d	n/a	n/a	n/a

Table 8

SEQ ID	PID NO.	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
109	7516593	1272419T6	SNP00000751	86	2122	A	G	A	n/coding	n/a	n/a	n/a
109	7516593	1358724H1	SNP00000751	256	2120	G	G	A	n/coding	n/a	n/a	n/a
109	7516593	2205012H1	SNP00062474	190	1321	C	C	T	Y421	n/a	n/a	n/a
109	7516593	2205012H1	SNP00068002	161	1292	A	A	G	S412	n/d	n/d	n/d
109	7516593	2395704H1	SNP00016249	94	1470	C	C	T	S471	n/a	n/a	n/a
109	7516593	2395704H1	SNP00068003	58	1434	C	C	T	T459	n/d	n/d	n/d
109	7516593	2395704H1	SNP00107662	178	1554	C	C	T	P499	n/d	n/a	n/a
109	7516593	343653T6	SNP00000751	65	2153	A	G	A	n/coding	n/a	n/a	n/a
109	7516593	6267143H1	SNP00050145	404	2032	C	T	C	n/coding	n/d	n/a	n/a
109	7516593	7380650H1	SNP00062474	122	1320	C	C	T	S421	n/a	n/a	n/a
109	7516593	7380650H1	SNP00068002	151	1291	A	A	G	V411	n/d	n/d	n/d
109	7516593	7380650H1	SNP00068003	9	1433	C	C	T	P459	n/d	n/d	n/d
110	7516603	033173H1	SNP00006647	119	466	T	T	C	A90	n/a	n/a	n/a
110	7516603	1371108H1	SNP00106653	41	358	C	C	A	Y54	n/a	n/a	n/a
110	7516603	1375926H1	SNP00074615	57	1070	C	C	A	n/coding	n/d	n/a	n/a
110	7516603	2757776H1	SNP0049719	200	619	C	C	T	T141	n/a	n/a	n/a
110	7516603	4256663T9	SNP00104109	503	343	T	C	T	N49	n/a	n/a	n/a
110	7516603	670023H1	SNP00153297	23	1095	C	C	T	n/coding	n/a	n/a	n/a
110	7516603	8626864J1	SNP00106653	238	375	C	C	A	A60	n/a	n/a	n/a
112	7525556	1402138F6	SNP00135928	23	704	T	T	C	D230	n/a	n/a	n/a
112	7525556	1424441H1	SNP00125721	143	161	C	C	A	L49	n/a	n/a	n/a
112	7525556	1710635T6	SNP00041515	137	1662	C	C	T	n/coding	n/a	n/a	n/a
112	7525556	2730123F6	SNP00041515	227	1661	C	C	T	n/coding	n/a	n/a	n/a
112	7525556	3660446F6	SNP00135927	176	673	A	A	G	H220	n/a	n/a	n/a
112	7525556	3660446H1	SNP00041514	139	638	T	T	C	A208	n/a	n/a	n/a
112	7525556	7686095H1	SNP00135927	552	674	G	A	G	Q220	n/a	n/a	n/a
112	7525556	7686095H1	SNP00135928	521	705	C	T	C	R231	n/a	n/a	n/a

What is claimed is:

1. An isolated polypeptide selected from the group consisting of:
 - a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56,
 - b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:5-6, SEQ ID NO:15, SEQ ID NO:17, SEQ ID NO:22, SEQ ID NO:25, SEQ ID NO:30-31, SEQ ID NO:33, SEQ ID NO:48, SEQ ID NO:52, and SEQ ID NO:54,
 - c) a polypeptide comprising a naturally occurring amino acid sequence at least 97% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1, SEQ ID NO:13, and SEQ ID NO:23,
 - d) a polypeptide comprising a naturally occurring amino acid sequence at least 93% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2, SEQ ID NO:21, and SEQ ID NO:34,
 - e) a polypeptide comprising a naturally occurring amino acid sequence at least 99% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:3, SEQ ID NO:32, SEQ ID NO:37, and SEQ ID NO:47,
 - f) a polypeptide comprising a naturally occurring amino acid sequence at least 91% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:7, and SEQ ID NO:19,
 - g) a polypeptide comprising a naturally occurring amino acid sequence at least 95% identical to the amino acid sequence of SEQ ID NO:9,
 - h) a polypeptide comprising a naturally occurring amino acid sequence at least 92% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:11 and SEQ ID NO:55,
 - i) a polypeptide comprising a naturally occurring amino acid sequence at least 98% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:28, SEQ ID NO:41, and SEQ ID NO:49,
 - j) a polypeptide comprising a naturally occurring amino acid sequence at least 96% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:29 and SEQ ID NO:53,
 - k) a polypeptide comprising a naturally occurring amino acid sequence at least 94% identical to the amino acid sequence of SEQ ID NO:42,
 - l) a polypeptide consisting essentially of a naturally occurring amino acid sequence at

least 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:4, SEQ ID NO:8, SEQ ID NO:10, SEQ ID NO:12, SEQ ID NO:14, SEQ ID NO:20, SEQ ID NO:24, SEQ ID NO:26-27, SEQ ID NO:35-36, SEQ ID NO:38-39, SEQ ID NO:43-46, SEQ ID NO:51, and SEQ ID NO:56,

5 m) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56, and

 n) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-56.

10 2. An isolated polypeptide of claim 1 comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-56. pf-1632p

 3. An isolated polynucleotide encoding a polypeptide of claim 1.

15 4. An isolated polynucleotide encoding a polypeptide of claim 2.

 5. An isolated polynucleotide of claim 4 comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:57-112.

20 6. A recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide of claim 3.

 7. A cell transformed with a recombinant polynucleotide of claim 6.

25 8. A transgenic organism comprising a recombinant polynucleotide of claim 6.

 9. A method of producing a polypeptide of claim 1, the method comprising:

 a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide, and said recombinant polynucleotide comprises a promoter sequence operably linked to a polynucleotide encoding the polypeptide of claim 1, and

30 b) recovering the polypeptide so expressed.

 10. A method of claim 9, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO:1-56.